Stochastic Optimization for Markov Modulated Networks with Application to Delay Constrained Wireless Scheduling

Michael J. Neely
University of Southern California
http://www-rcf.usc.edu/∼mjneely

***Abstract*— We consider a wireless system with a small number of delay constrained users and a larger number of users without delay constraints. We develop a scheduling algorithm that reacts to time varying channels and maximizes throughput utility (to within a desired proximity), stabilizes all queues, and satisfies the delay constraints. The problem is solved by reducing the constrained optimization to a set of weighted stochastic shortest path problems, which act as natural generalizations of max-weight policies to Markov modulated networks. We also present approximation results that do not require a-priori statistical knowledge, and discuss the additional complexity and delay incurred as compared to systems without delay constraints. The solution technique is general and applies to other constrained stochastic network optimization problems.**

I. Introduction

This paper considers delay-aware scheduling in a multi-user wireless uplink or downlink with K delay-constrained users and N delay-unconstrained users, each with different transmission channels. The system operates in slotted time with normalized slots $t \in \{0, 1, 2, \ldots\}$. Every slot, a random number of new packets arrive from each user. Packets are queued for eventual transmission, and every slot a scheduler looks at the queue backlog and the current channel states and chooses one channel to serve. The number of packets that are transmitted over that channel depends on its current channel state. The goal is to stabilize all queues, satisfy average delay constraints for the delay-constrained users, and drop as few packets as possible.

Without the delay constraints, this problem is a classical *opportunistic scheduling* problem, and can be solved with efficient max-weight algorithms based on Lyapunov drift and Lyapunov optimization (see [1] and references therein). The delay constraints make the problem a much more complex *Markov Decision Problem* (MDP). While general methods for solving MDPs exist (see, for example, [2] [3] [4] [5]), they typically suffer from a curse of dimensionality. Specifically, the number of queue state vectors grows geometrically in the number of queues. Thus, a general problem with many queues has an intractably large state space. This creates non-polynomial implementation complexity for offline approaches such as linear programming [3] [4], and non-polynomial complexity and/or learning time for online or quasi online/offline approaches such as Q-learning [2] [6].

We do not solve this fundamental curse of dimensionality. Rather, we avoid this difficulty by focusing on the special structure that arises in a wireless network with a *relatively small number of delay-constrained users* (say, $K \leq 5$), but with an arbitrarily large number of users without delay constraints (so that N can be large). This is an important scenario, particularly in cases when the number of “best effort” users in a network is much larger than the number of delay-constrained users. We develop a solution that, on each slot, requires a computation that has a complexity that depends geometrically in K, but only polynomially in N. Further, the resulting convergence times and delays are fully polynomial in the total number of queues $K+N$. Our solution uses a concept of *forced renewals* that introduces a deviation from optimality that can be made arbitrarily small with a corresponding polynomial tradeoff in convergence time. Finally, we show that a simple Robbins-Monro approximation technique can be used, without knowledge of the channel or traffic statistics, and yields similar performance. Our methods are general and can be applied to other MDPs for queueing networks with similar structure.

Related prior work on delay optimality for multi-user opportunistic scheduling under special symmetric assumptions is developed in [7] [8] [9], and single-queue delay optimization problems are treated in [10] [11] [12] [13] using dynamic programming and Markov Decision theory. Optimal asymptotic energy-delay tradeoffs are developed for single queue systems in [14], and optimal energy-delay and utility-delay tradeoffs for multi-queue systems are treated in [15] [16]. The algorithms of [15] [16] have very low complexity and converge quickly even for large networks, although the tradeoff-optimal delay guarantees they achieve do not necessarily optimize the coefficient multiplier in the delay expression.

Our approach in the present paper treats the MDP problem associated with delay constraints using Lyapunov drift and Lyapunov optimization theory [1]. We extend the max-weight principles for stochastic network optimization to treat *Markov-modulated networks*, where the network costs depend on both the control actions taken and the current state (such as

Michael J. Neely is with the Electrical Engineering department at the University of Southern California, Los Angles, CA (web: http://www-rcf.usc.edu/∼mjneely).

This material is supported in part by one or more of the following: the DARPA IT-MANET program grant W911NF-07-0028, the NSF grant OCE 0520324, the NSF Career grant CCF-0747525.

the queue state) the system is in. For each cost constraint we define a *virtual queue*, and show that the constrained MDP can be solved using Lyapunov drift theory implemented over a variable-length frame, where "max-weight" rules are replaced with weighted stochastic shortest path problems. This is similar to the Lagrange multiplier approaches used in the related works [12] [13] that treat power minimization for single-queue wireless links with an average delay constraint. The work in [12] uses stochastic approximation with a 2-timescale argument and a limiting ordinary differential equation (ODE). The work in [13] treats a single-queue MIMO system using primal-dual updates [17]. Our virtual queues are similar to the Lagrange Multiplier updates in [12] [13]. However, we treat multi-queue systems, and we use a different analytical approach that emphasizes stochastic shortest paths over variable length frames. Our resulting algorithm has an implementation complexity that grows geometrically in the number of delay-constrained queues K, but polynomially in the number of delay-unconstrained queues N. Further, we obtain polynomial bounds on convergence times and delays.

The next section describes the general stochastic network model and its application to delay constrained wireless systems. Section III presents the weighted shortest-path algorithm. Section IV describes an approximate implementation that does not require a-priori knowledge of channel or traffic probabilities. The approximation scheme learns by observing past system outcomes, and uses a classic Robbins-Monro iteration (see [2]) together with a *delayed queue analysis* to uncorrelate past samples from current queue states. Section V treats a more general problem of optimizing convex functions of time average penalties.

II. Network Model

Consider the following abstract model of a stochastic queueing network (application to delay constrained wireless systems is detailed in Section II-C). The system operates in slotted time $t \in \{0, 1, 2, \ldots\}$. Let $\mathcal{N}$ represent a finite set of queues to be stabilized, and let $\boldsymbol{Q}(t) = (Q_n(t))_{n\in\mathcal{N}}$ denote the vector of queue backlogs on slot t. Each queue $Q_n(t)$ is assumed to have infinite buffer space, and has dynamic update equation:

$$Q_n(t+1) = \max[Q_n(t) - \mu_n(t), 0] + R_n(t) \tag{1}$$

where $\mu_n(t)$ and $R_n(t)$ are the service rate and new arrivals, respectively, for queue n on slot t. Let $\boldsymbol{\mu}(t)$ and $\boldsymbol{R}(t)$ be the vector of service rates and arrival variables with entries $n \in \mathcal{N}$. On each slot t, the vectors $\boldsymbol{\mu}(t)$ and $\boldsymbol{R}(t)$ are determined as functions of a *random outcome* $\Omega(t)$, a *state variable* $z(t)$, and a *control action* $I(t)$:

$$\begin{aligned} \boldsymbol{\mu}(t) &\triangleq \hat{\boldsymbol{\mu}}(I(t), \Omega(t), z(t)) \\ \boldsymbol{R}(t) &\triangleq \hat{\boldsymbol{R}}(I(t), \Omega(t), z(t)) \end{aligned}$$

Specifically, the control action $I(t)$ is made every slot with knowledge of $z(t)$ and $\Omega(t)$ (and also $\boldsymbol{Q}(t)$), and is constrained to take values in an abstract set $\mathcal{I}_{\Omega(t),z(t)}$ that has arbitrary cardinality and that possibly depends on $\Omega(t)$ and $z(t)$. The random outcome $\Omega(t)$ takes values in a set with arbitrary cardinality, and represents a collection of network parameters (such as channel states) that can randomly change from slot to slot. We assume that $\Omega(t)$ is i.i.d. over slots with some fixed (but potentially unknown) distribution that does not depend on the current state or the past network control actions. The state variable $z(t)$ takes values in a finite or countably infinite set $\mathcal{Z}$, and represents a controlled Markov chain related to the network (this will be used to represent delay-constrained queues in the next subsection). The transition probabilities of $z(t)$ depend on $\Omega(t)$ and on the control decision $I(t)$. That is, for all states $y, z \in \mathcal{Z}$, we define $P_{yz}(I, \Omega)$ as follows:

$$P_{yz}(I, \Omega) \triangleq Pr[z(t+1) = z \mid z(t) = y, I(t) = I, \Omega(t) = \Omega]$$

The state space $\mathcal{Z}$ is assumed to contain a state 0 that is accessible from any state $z \in \mathcal{Z}$. In the next sub-section, we impose an additional *ϕ-forced renewal assumption*, where the probability of reaching state 0 from any state $z \in \mathcal{Z}$ and under any $\Omega(t)$, $I(t)$ is at least ϕ, for some positive probability $\phi > 0$ (described in more detail in Section II-B).

For each slot t we have a collection of general *network penalties* $x_m(t)$ for $m \in \{0, 1, \ldots, M\}$ for some finite integer M. These are defined by *penalty functions* $\hat{x}_m(\cdot)$ that represent different types of costs incurred when a control action $I(t)$ is taken under outcome $\Omega(t)$ and state variable $z(t)$:

$$x_m(t) \triangleq \hat{x}_m(I(t), \Omega(t), z(t))$$

The penalty functions are arbitrary and possibly negative (negative penalties can represent rewards). However, they are assumed to be upper and lower bounded by finite constants x_m^{min} and x_m^{max}, so that regardless of $I(t)$, $\Omega(t)$, $z(t)$ we have:

$$x_m^{min} \leq \hat{x}_m(\cdot) \leq x_m^{max}$$

Similarly, the $\hat{\boldsymbol{\mu}}(\cdot)$ and $\hat{\boldsymbol{R}}(\cdot)$ functions are arbitrary but are assumed to be bounded by constants μ_n^{max} and R_n^{max}:

$$0 \leq \hat{\mu}_n(\cdot) \leq \mu_n^{max} \ \ , \ \ 0 \leq \hat{R}_n(\cdot) \leq R_n^{max} \ \ \forall n \in \mathcal{N}$$

For each penalty $m \in \{0, 1, \ldots, M\}$, each queue $Q_n(t)$ for $n \in \mathcal{N}$, and for a given control policy that makes decisions $I(t)$ over time, we define the following time averages:

$$\begin{aligned} \overline{x}_m &\triangleq \limsup_{t\rightarrow\infty} \frac{1}{t}\sum_{\tau=0}^{t-1} \mathbb{E}\left\{\hat{x}_m(I(\tau), \Omega(\tau), z(\tau))\right\} \\ \overline{Q}_n &\triangleq \limsup_{t\rightarrow\infty} \frac{1}{t}\sum_{\tau=0}^{t-1} \mathbb{E}\left\{Q_n(\tau)\right\} \end{aligned}$$

We say that a queue $Q_n(t)$ is *stable* if $\overline{Q}_n < \infty$.[1] We now state the stochastic optimization problem of interest.

Stochastic Optimization Problem: Determine a control policy that solves:

$$\begin{aligned} &\text{Minimize:} && \overline{x}_0 && (2) \\ &\text{Subject to:} && \overline{x}_m \leq x_m^{av} \ \text{ for all } m \in \{1, \ldots, M\} && (3) \\ & && \overline{Q}_n < \infty \ \text{ for all } n \in \mathcal{N} && (4) \end{aligned}$$

where each constant x_m^{av} represents a desired constraint on the time average of penalty $x_m(t)$.

[1] This is often called *strongly stable* as it implies finite average queue backlog.

The stochastic optimization problem seeks to minimize the time average of penalty $x_0(t)$ subject to constraints on the time averages of all other penalties, and on the stability of all queues $Q_n(t)$ for $n \in \mathcal{N}$. This is similar to the stochastic network optimization problems of [1] [18] [19], with the exception that the penalty function now includes the state variable $z(t) \in \mathcal{Z}$, and the transition probabilities for $z(t)$ depend on the control action $I(t)$ and the random i.i.d. variable $\Omega(t)$. We say that the problem is a *stochastic feasibility problem* if we desire only to satisfy the time average constraints (3)-(4), without regard for minimization of $\overline{x}_0$. The problem (2)-(4) is generalized in Section V to treat optimization of convex functions of time averages, similar to the objectives considered without the Markov modulated state variable $z(t)$ in [1] [20] [21].

A. Existence of a Maximizer for Linear Functionals

We have assumed that the control action $I(t)$ takes values in an abstract set $\mathcal{I}_{\Omega(t),z(t)}$ with arbitrary cardinality, and that the probability transition matrix $P_{yz}(I,\Omega)$ is an arbitrary function of I, and the corresponding functions $\hat{\mu}_n(I(t),\Omega(t),z(t))$, $\hat{R}_n(I(t),\Omega(t),z(t))$, $\hat{x}_m(I(t),\Omega(t),z(t))$ are arbitrary but bounded. It turns out that our resulting control algorithm will involve choosing $I(t)$ to maximize a weighted sum of these functions every slot. Hence, it is useful to assume throughout the paper that for any given Ω, z, and for any (possibly negative) scalars $\{\alpha_n, \beta_n, \gamma_m, \delta_y\}$, problems of the type:

$$\begin{aligned} \text{Maximize:} \quad & \textstyle\sum_n \alpha_n \hat{\mu}_n(I,\Omega,z) + \sum_n \beta_n \hat{R}_n(I,\Omega,z) \\ & + \textstyle\sum_m \gamma_m \hat{x}_m(I,\Omega,z) + \sum_y P_{zy}(I,\Omega)\delta_y \\ \text{Subject to:} \quad & I \in \mathcal{I}_{\Omega,z} \end{aligned}$$

have at least one well defined maximizer $I^* \in \mathcal{I}_{\Omega,z}$. This holds in most practical situations, such as whenever the set $\mathcal{I}_{\Omega,z}$ is finite for all Ω and z. Alternatively, it also holds whenever $\mathcal{I}_{\Omega,z}$ is a compact subset of $\mathbb{R}^L$ (for some finite integer L), the state space $\mathcal{Z}$ is finite (so that the sum $\sum_y P_{zy}(\cdot)$ has a finite number of terms), and the functions $\hat{\mu}_n(\cdot)$, $\hat{R}_n(\cdot)$, $\hat{x}_m(\cdot)$, $P_{zy}(\cdot)$ are continuous in I for all Ω and z.

B. The Forced Renewal Assumption

To ensure that the $z(t)$ state variable "renews" itself regularly by returning to the 0 state, we consider the following simple (and sub-optimal) mechanism. Let $\Omega(t) \triangleq [\omega(t);\phi(t)]$, where $\omega(t)$ is the random outcome of network state variables (taking values in an abstract set $\mathcal{W}$ with arbitrary cardinality), and $\phi(t)$ is an independent Bernoulli 0/1 variable that is i.i.d. over slots with $Pr[\phi(t)=1]=\phi$, for some small but positive *forced renewal probability* $\phi > 0$. If $\phi(t)=1$, the system experiences a *forced renewal event* which ensures that $z(t+1)=0$. Thus, the transition probabilities have the following property for all $\omega \in \mathcal{W}$, $y \in \mathcal{Z}$, and $I \in \mathcal{I}_{[\omega;1],y}$:

$$\begin{aligned} P_{yz}(I,\Omega) = 0 & \quad \text{if } \Omega = (\omega,1) \text{ and } z \neq 0 \\ P_{y0}(I,\Omega) = 1 & \quad \text{if } \Omega = (\omega,1) \end{aligned}$$

The value of $\phi(t)$ is known to the network controller at the beginning of slot t, although if $\phi(t)=1$ the renewal itself only occurs at the end of slot t when the next state $z(t+1)$ is forced to 0. In this way, the control action taken during a slot t in which $\phi(t)=1$ still affects the queueing and penalty functions $\hat{\boldsymbol{\mu}}(I(t),\Omega(t),z(t))$, $\hat{\boldsymbol{R}}(I(t),\Omega(t),z(t))$, $\hat{x}_m(I(t),\Omega(t),z(t))$, and these functions possibly have different values when $\phi(t)=0$ versus $\phi(t)=1$ (recall that $\Omega(t)=[\omega(t);\phi(t)]$, so that the functions can also depend on $\phi(t)$).

This forced renewal structure implicitly assumes that the system can physically reset the state variable $z(t)$ to zero on any slot. Further, even if the system has this physical capability, it is generally sub-optimal to force such renewals with probability ϕ every slot. However, for many systems of interest (such as the network defined in the next subsection), if ϕ is small then the optimal performance over systems constrained by this *ϕ-forced renewal assumption* is close to the optimal performance for systems without forced renewals. Throughout this paper, we define optimality in terms of systems with ϕ-forced renewals, with the understanding that ϕ is a small but positive value.

C. Wireless Systems with Delay Constraints

Consider now the following wireless system that operates in discrete time and fits the abstract model defined above. Let $\mathcal{N}\triangleq\{1,\ldots,N\}$ denote a set of delay-unconstrained queues, and let $\mathcal{K}\triangleq\{1,\ldots,K\}$ represent a set of delay-constrained queues. All packets have fixed lengths, and we let $\boldsymbol{Q}(t) = (Q_n(t))|_{n\in\mathcal{N}}$ and $\boldsymbol{Z}(t) = (Z_k(t))|_{k\in\mathcal{K}}$ be the vector of integer queue lengths in all delay-unconstrained and delay-constrained queues, respectively, on slot t. Suppose each delay-constrained queue has a finite buffer size B_{max}, and let $\mathcal{Z}$ represent the state space for $\boldsymbol{Z}(t)$, which has a finite size of $(B_{max}+1)^K$. To emphasize membership in the state space $\mathcal{Z}$ (and to simplify notation for the transition probabilities), we let $z(t)$ represent the vector state $\boldsymbol{Z}(t)$.

Forced renewals occur according to the i.i.d. Bernoulli process $\phi(t)$ with forced renewal probability $\phi > 0$. If a forced renewal occurs on slot t (so that $\phi(t)=1$), all data in all delay-constrained queues $k \in \mathcal{K}$ is dropped at the end of the slot, so that $z(t+1)=0$ (where the state $0 \in \mathcal{Z}$ represents the vector of all zeros). The data in the delay-unconstrained queues is not dropped. The maximum drop rate in a queue $k \in \mathcal{K}$ due to such forced renewals is at most $(B_{max}+\lambda_k)\phi$ drops/slot, where λ_k is the rate of new arrivals to queue k. The value $(B_{max}+\lambda_k)\phi$ can be made arbitrarily small with a small choice of ϕ.

Let $A_i(t)$ represent the number of new packet arrivals for user $i \in \mathcal{N}\cup\mathcal{K}$ on slot t. Let $S_i(t)$ be the current channel state for user $i \in \mathcal{N}\cup\mathcal{K}$. Specifically, $S_i(t)$ is a non-negative integer that represents the number of packets that can be transmitted over channel i on slot t if the channel is selected for transmission. Let $\boldsymbol{A}(t)$ and $\boldsymbol{S}(t)$ be vectors of $A_i(t)$ and $S_i(t)$ components. Assume that the joint vector $[\boldsymbol{A}(t),\boldsymbol{S}(t)]$ is i.i.d. over slots (possibly with correlated entries). Arrivals and channels are assumed to be bounded by constants A_{max}, S_{max}, so that $A_i(t) \leq A_{max}$ and $S_i(t) \leq S_{max}$ for all i and t. Every slot the controller observes all channel states and must select a single channel (either from the delay-constrained or delay-unconstrained queues) to serve.

For each finite buffer queue $k \in \mathcal{K}$, the controller makes an additional admit/drop decision immediately upon packet arrival (subject to the finite buffer constraint). Let $R_k(t)$ and $D_k(t)$ respectively represent the amount of new arrivals added on slot t and new packets dropped on slot t, where:

$$R_k(t) + D_k(t) = A_k(t) \tag{5}$$

The queue update equation is given by:

$$Z_k(t+1) = \begin{cases} \max[Z_k(t) - \mu_k(t), 0] + R_k(t) & \text{if } \phi(t) = 0 \\ 0 & \text{if } \phi(t) = 1 \end{cases} \tag{6}$$

where for each $i \in \mathcal{N} \cup \mathcal{K}$ we have $\mu_i(t) = S_i(t)$ if channel i is served on slot t, and $\mu_i(t) = 0$ else.

Let $\omega(t) \triangleq [\boldsymbol{A}(t), \boldsymbol{S}(t)]$ be a combined system state variable that captures the random arrivals and channels, and let $\Omega(t) \triangleq [\omega(t), \phi(t)]$. Let $I(t)$ be a combined *control action*, which indicates which channel $i \in \mathcal{K} \cup \mathcal{N}$ to serve, and how to choose the admit/drop variables $R_k(t)$ and $D_k(t)$ for $k \in \mathcal{K}$. The control action $I(t)$ is made with knowledge of $\Omega(t)$ and $z(t)$ (and also $\boldsymbol{Q}(t)$). The constraint set $\mathcal{I}_{\Omega(t),z(t)}$ is defined to ensure that at most one channel $i \in \mathcal{K} \cup \mathcal{N}$ is served, and that packet drops act according to (5)-(6) and satisfy the finite buffer constraint $Z_k(t+1) \leq B_{max}$. Given the $\Omega(t)$, $z(t)$, and $I(t)$ values, the next-state is deterministically known, so that the transition probabilities $P_{zy}(I, \Omega)$ are either 0 or 1. Finally, the queueing dynamics for each delay-unconstrained queue $n \in \mathcal{N}$ are given by (1), with arrival and service functions given by:

$$\hat{R}_n(I(t), \Omega(t), z(t)) = A_n(t)$$

$$\hat{\mu}_n(I(t), \Omega(t), z(t)) = \begin{cases} S_n(t) & \text{if } n \text{ is served on slot } t \\ 0 & \text{otherwise} \end{cases}$$

Note in this case that the input to each delay-unconstrained queue $n \in \mathcal{N}$ is the (uncontrolled) random process $A_n(t)$.

D. Example Penalties for Average Congestion and Delay

To use the framework of abstract penalty functions to enforce an average congestion bound on queue $Z_k(t)$ (for a given $k \in \mathcal{K}$), we can define a penalty function of the form:

$$\hat{x}_k(I(t), \Omega(t), z(t)) = Z_k(t)$$

This penalty function does not use the $I(t)$ or $\Omega(t)$ arguments, and uses the fact that $Z_k(t)$ is a component of the $z(t)$ state variable. Enforcing a constraint of the type $\overline{x}_k \leq x_k^{av}$ ensures that average queue congestion is no more than x_k^{av}.

To enforce a constraint on the time average rate of dropping packets in a delay-constrained queue $k \in \mathcal{K}$, we can define a penalty function of the form:

$$\hat{x}_k(I(t), \Omega(t), z(t)) = \begin{cases} A_k(t) - R_k(t) & \text{if } \phi(t) = 0 \\ A_k(t) + Z_k(t) - \tilde{\mu}_k(t) & \text{if } \phi(t) = 1 \end{cases}$$

where $\tilde{\mu}_k(t) \triangleq \min[\mu_k(t), Z_k(t)]$ and represents the number of packets served in queue k on slot t. In this case, the penalty is equal to the exact amount of packet drops in queue k on slot t, so that ensuring $\overline{x}_k \leq x_k^{av}$ enforces a constraint on the time average rate of packet drops. Defining a penalty function $x_0(\cdot)$ as a (possibly weighted) sum of packet drops in all of the delay-constrained queues $k \in \mathcal{K}$ allows for minimization of a weighted sum of packet drop rates subject to additional desired constraints.

Finally, to enforce a constraint that the average delay of (non-dropped) packets in a queue $k \in \mathcal{K}$ is less than or equal to some desired bound W_k^{av} (where W_k^{av} is a given constant), we can use a penalty function of the form:

$$\hat{x}_k(I(t), \Omega(t), z(t)) = Z_k(t) - \tilde{\mu}_k(t) W_k^{av}$$

and enforce the constraint $\overline{x}_k \leq 0$. Assuming time average limits are well defined, this ensures that

$$\overline{Z}_k - \tilde{\lambda}_k W_k^{av} \leq 0 \tag{7}$$

where $\tilde{\lambda}_k$ is the time average rate of actual packets served in queue k (and is also the time average rate of non-dropped packets that are admitted). By Little's Theorem [22], we have:

$$\overline{Z}_k = \tilde{\lambda}_k \overline{W}_k$$

where $\overline{W}_k$ is the average delay of non-dropped packets in queue k, and so from (7) we deduce that $\overline{W}_k \leq W_k^{av}$ (assuming that $\tilde{\lambda}_k > 0$).

E. Slackness Assumptions

Consider the general stochastic queueing network model, and let $\mathcal{M} \triangleq \{1, \ldots, M\}$ represent the set of penalties involved in the feasibility constraints (3)-(4).

Definition 1: A control policy $I(t)$ is a (z, Ω)*-only policy* if it satisfies the ϕ-forced renewal assumption and it makes stationary and possibly randomized control actions $I(t) \in \mathcal{I}_{\Omega(t),z(t)}$ for each slot t based only on the current $\Omega(t)$ and $z(t)$ (and hence independently of $\boldsymbol{Q}(t)$).

Suppose there exists a (z, Ω)-only policy $I^*(t)$ that satisfies the feasibility constraints (3)-(4). Let $z^*(t)$ represent the resulting network state variable under this policy, and note it evolves according to an irreducible finite or countably infinite state Markov chain. Hence, time average limits are well defined [22]. Let $\overline{x}_m^*$, $\overline{\mu}_n^*$, $\overline{r}_n^*$ respectively represent the time average of penalty $x_m(t)$, transmission $\mu_n(t)$, and admission $A_n(t)$, under policy $I^*(t)$. Because queue stability requires the time average arrival rate to be less than or equal to the time average service rate, it is easy to show that (3)-(4) imply:

$$\overline{x}_m^* \leq x_m^{av} \quad \text{for all } m \in \mathcal{M} \tag{8}$$

$$\overline{\mu}_n^* - \overline{r}_n^* \geq 0 \quad \text{for all } n \in \mathcal{N} \tag{9}$$

Let x_0^{opt} represent the infimum value of $\overline{x}_0$ over all (z, Ω)-only policies that satisfy (8)-(9). We shall measure optimality of our algorithm designs with respect to x_0^{opt}. This is typically non-restrictive. For example, if $\mathcal{Z}$ has a finite state space, it can be shown that the infimum of $\overline{x}_0$ over *all policies* that satisfy (3)-(4) is equal to x_0^{opt}.[2]

Assume that $z(0) = 0$, and define *renewal events* as times $\{t_g\}_{g=0}^{\infty}$, starting with $t_0 = 0$, where each renewal event t_g occurs when $z(t_g) = 0$ and some other criterion is met (as

[2] This can be shown by well known optimality of stationary randomized policies for MDP problems over finite state spaces [4] and for queue stability problems [18], although the formal proof is omitted for brevity.

described below). Define a *renewal interval* as the duration of time between successive renewal events (including the starting renewal slot but not including the ending renewal slot). Define T_g as the size of the gth renewal interval (also called the *inter-renewal time*). The additional criterion that defines a renewal can be anything that satisfies the following *renewal requirements*:

- There are finite constants m_1 and m_2 such that under any policy for choosing $I(t)$ over time, the inter-renewal times have first and second moments upper bounded by m_1 and m_2, respectively, regardless of past history.
- Under any (z, Ω)-only policy $I^*(t)$, the inter-renewal times are independent and identically distributed (i.i.d.), as are the sequences of decisions and penalties incurred over different renewal intervals.

Note that these requirements are met whenever the system "resets" itself at renewal events, so that under a particular (z, Ω)-only policy the system has independent but identically distributed behavior on each renewal interval. By basic renewal theory, all time average penalties have well defined limits that are exactly equal to the expected sum penalty over a renewal interval divided by the expected duration of the renewal interval [23].

For our purposes, we focus on the following three different examples of renewal definitions. Systems with "type-1 renewals" have renewals defined by any slot t_g at which $z(t_g) = 0$. Note that a type-1 renewal may arise either because of a *forced renewal*, or because of controller decisions that lead to the $z(t) = 0$ state. Hence, the average duration of a type-1 renewal interval is less than or equal to $1/\phi$. Alternatively, "type-2 renewals" are defined only by forced renewal events, and hence have average size exactly equal to $1/\phi$. Finally, "type-3 renewals" are defined by every bth visitation to the $z(t) = 0$ state, where b is a given positive integer. Thus, the average duration of a type-3 renewal is less than or equal to b/ϕ. Note that all three definitions meet the renewal requirements specified above.

Consider any valid renewal definition for the network (such as a type-1, type-2, or type-3 definition). Suppose that $z(0) = 0$ and that time 0 is a renewal time. Define T^* as the random time until the next renewal event under the (z, Ω)-only policy $I^*(t)$ that satisfies (8)-(9). We have by basic renewal theory:

$$\overline{x}_m^* = \frac{\mathbb{E}\left\{\sum_{\tau=0}^{T^*-1} \hat{x}_m(I^*(\tau), \Omega(\tau), z^*(\tau))\right\}}{\mathbb{E}\left\{T^*\right\}} \quad \forall m \in \mathcal{M}$$

$$\overline{\mu}_n^* - \overline{r}_n^* = \frac{\mathbb{E}\left\{\sum_{\tau=0}^{T^*-1} \hat{d}_n(I^*(\tau), \Omega(\tau), z^*(\tau))\right\}}{\mathbb{E}\left\{T^*\right\}} \quad \forall n \in \mathcal{N}$$

where $\hat{d}_n(I, \Omega, z)$ is defined:

$$\hat{d}_n(I, \Omega, z) \triangleq \hat{\mu}_n(I, \Omega, z) - \hat{R}_n(I, \Omega, z)$$

In addition to assuming the feasibility constraints (8)-(9) are satisfied, we make the following two mild assumptions. The first is a *slackness assumption* that is a stochastic analogue of a *Slater condition* for static optimization problems [17].

Assumption 1: (Slackness of Feasibility) There exists a value $\epsilon > 0$ such that the constraints of (8)-(9) can be met with ϵ slackness. Specifically, there exists a (z, Ω)-only policy $I^*(t)$ that satisfies the following for all $m \in \mathcal{M}$ and $n \in \mathcal{N}$:

$$\frac{\mathbb{E}\left\{\sum_{\tau=0}^{T^*-1} x_m^*(\tau)\right\}}{\mathbb{E}\left\{T^*\right\}} \leq x_m^{av} - \epsilon \tag{10}$$

$$\frac{\mathbb{E}\left\{\sum_{\tau=0}^{T^*-1} d_n^*(\tau)\right\}}{\mathbb{E}\left\{T^*\right\}} \geq \epsilon \tag{11}$$

where for notational simplicity we have defined:

$$\begin{aligned} x_m^*(\tau) &\triangleq \hat{x}_m(I^*(\tau), \Omega(\tau), z^*(\tau)) \\ d_n^*(\tau) &\triangleq \hat{d}_n(I^*(\tau), \Omega(\tau), z^*(\tau)) \end{aligned}$$

Assumption 2: (Optimization) There exists a (z, Ω)-only policy $I^*(t)$ (not necessarily the same policy as in Assumption 1) that satisfies:

$$\frac{\mathbb{E}\left\{\sum_{\tau=0}^{T^*-1} x_0^*(\tau)\right\}}{\mathbb{E}\left\{T^*\right\}} = x_0^{opt} \tag{12}$$

$$\frac{\mathbb{E}\left\{\sum_{\tau=0}^{T^*-1} x_m^*(\tau)\right\}}{\mathbb{E}\left\{T^*\right\}} \leq x_m^{av} \quad \forall m \in \mathcal{M} \tag{13}$$

$$\frac{\mathbb{E}\left\{\sum_{\tau=0}^{T^*-1} d_n^*(\tau)\right\}}{\mathbb{E}\left\{T^*\right\}} \geq 0 \quad \forall n \in \mathcal{N} \tag{14}$$

where T^* is the size of the first renewal interval, and $z^*(\tau)$ is the network state at time τ under policy $I^*(t)$.

Assumption 1 states that there is a (z, Ω)-only policy that satisfies all feasibility constraints (and queue stability constraints) with ϵ slackness. Assumption 2 states that there is another, typically different, (z, Ω)-only policy that achieves the desired optimum value x_0^{opt} while satisfying the feasibility constraints (possibly with no slackness in these constraints).[3]

III. THE DYNAMIC CONTROL ALGORITHM

To solve the stochastic feasibility and stochastic optimization problems for our queueing network, we extend the framework of [1] to a case of variable length frames. Specifically, for each time average penalty constraint (3), parameterized by $m \in \mathcal{M} \triangleq \{1, \ldots, M\}$, we define a *virtual queue* $Y_m(t)$ that is initialized to zero and has dynamic update equation:

$$Y_m(t+1) = \max[Y_m(t) - x_m^{av} + x_m(t), 0] \tag{15}$$

where $x_m(t) = \hat{x}_m(I(t), \Omega(t), z(t))$ is the penalty incurred on slot t by a particular choice of the control decision $I(t)$ (under the observed $\Omega(t)$ and $z(t)$). The intuition is that if the virtual queue $Y_m(t)$ is stable, then the time average rate of the "input process" $x_m(t)$ is less than or equal to the "service rate" x_m^{av} [18]. This turns the time average constraint into a simple queue stability problem.[4]

[3] In the case when the infimum value x_0^{opt} is only achievable over a limit of an infinite sequence of (z, Ω)-only policies, we can replace x_0^{opt} with $x_0^{achieve}$, where $x_0^{achieve}$ is any achievable value, and then recover the results of Theorem 2 by taking a limit as $x_0^{achieve} \rightarrow x_0^{opt}$.

[4] Note that $Y_m(t)$ can be viewed as a "generalized" queue, as the "service rate" x_m^{av} can be negative, as can the $x_m(t)$ value.

A. Lyapunov Drift

Define $\boldsymbol{Y}(t)$ as a vector of all virtual queues $Y_m(t)$ for $m \in \mathcal{M}$, and define $\boldsymbol{\Theta}(t) \triangleq [\boldsymbol{Y}(t); \boldsymbol{Q}(t)]$ as the combined queue vector. Assume all queues are initially empty, so that $\boldsymbol{\Theta}(0) = \boldsymbol{0}$. Define the following quadratic Lyapunov function:

$$L(\boldsymbol{\Theta}(t)) \triangleq \frac{1}{2}\sum_{n\in\mathcal{N}} Q_n(t)^2 + \frac{1}{2}\sum_{m\in\mathcal{M}} Y_m(t)^2$$

Suppose time t_g is a renewal event (for any valid renewal definition), and let T be the random time until the next renewal event (which may depend on the control policy, such as when type-1 renewals are used). Define the *variable-slot conditional Lyapunov drift* $\Delta_T(\boldsymbol{\Theta}(t_g))$ as follows:[5]

$$\begin{aligned}&\Delta_T(\boldsymbol{\Theta}(t_g)) \triangleq \\ &\qquad \mathbb{E}\left\{L(\boldsymbol{\Theta}(t_g+T)) - L(\boldsymbol{\Theta}(t_g)) \,|\, \boldsymbol{\Theta}(t_g), z(t_g)=0\right\} \quad (16)\end{aligned}$$

The expectation in the drift definition above is with respect to the random renewal interval duration T, the random events that can take place over this interval, and the possibly random control actions $I(t)$ that are made during this interval. The explicit conditioning on $z(t_g) = 0$ in (16) will be suppressed in the remainder of this paper, as this conditioning is implied given that t_g is a renewal time.

It is important to note the following subtlety: The renewal events under a given policy $I(t)$ may arise from the decisions made under the policy (as in type-1 or type-3 definitions), although the implemented policy $I(t)$ may not be stationary and/or may depend on the queue values $\boldsymbol{Q}(t)$, and so actual system events are not necessarily i.i.d. over different renewals. Therefore, these "renewal-events" do not necessarily reset the system dynamics of the actual system. However, these times act as convenient "time-stamps" over which to analytically compare the Lyapunov drift of the actual implemented policy with the corresponding drifts of the (z, Ω)-only policies of Assumptions 1 and 2.

Lemma 1: (Lyapunov Drift) Under any network control policy for choosing $I(t)$ over time, and for any renewal definition that meets the renewal requirements, the variable-slot conditional Lyapunov drift satisfies the following at any renewal time t_g and any $\boldsymbol{\Theta}(t_g)$:

$$\Delta_T(\boldsymbol{\Theta}(t_g)) \quad \leq \quad B + D(\boldsymbol{\Theta}(t_g)) \qquad (17)$$

where $D(\boldsymbol{\Theta}(t_g))$ is defined:

$$\begin{aligned}D(\boldsymbol{\Theta}(t_g)) \triangleq &-\textstyle\sum_{n\in\mathcal{N}} Q_n(t_g)\mathbb{E}\left\{\sum_{\tau=0}^{T-1} d_n(t_g+\tau) \,|\, \boldsymbol{\Theta}(t_g)\right\} \\ &-\textstyle\sum_{m\in\mathcal{M}} Y_m(t_g)\mathbb{E}\left\{T x_m^{av} - \sum_{\tau=0}^{T-1} x_m(t_g+\tau) \,|\, \boldsymbol{\Theta}(t_g)\right\} (18)\end{aligned}$$

where we recall that:

$$d_n(t) \triangleq \hat{d}_n(I(t), \Omega(t), z(t)) \ \ , \ \ x_m(t) \triangleq \hat{x}_m(I(t), \Omega(t), z(t))$$

and where B is a finite constant defined:

$$B \triangleq \frac{\sigma^2}{2} \sup_{\boldsymbol{\Theta}(t_g)} \mathbb{E}\left\{T^2 \,|\, \boldsymbol{\Theta}(t_g)\right\}$$

where σ^2 is a constant that satisfies the following for all t:

$$\sigma^2 \geq \sum_{n\in\mathcal{N}} [\mu_n(t)^2 + R_n(t)^2] + \sum_{m\in\mathcal{M}} (x_m(t) - x_m^{av})^2 \qquad (19)$$

Note that σ^2 is finite due to the finite bounds on the penalties $x_m(t)$ and on the queue variables $\mu_n(t)$ and $R_n(t)$, and B is finite and bounded by $m_2\sigma^2/2$ by the second moment property of the renewal requirements in Section II-E. Further, for type-1 or type-2 renewals, the second moment of T is bounded by the corresponding second moment of a geometric random variable (with success probability ϕ), so that for all $\boldsymbol{\Theta}(t_g)$:

$$\mathbb{E}\left\{T^2 \,|\, \boldsymbol{\Theta}(t_g)\right\} \leq (2-\phi)/\phi^2$$

For type-3 renewals the second moment is bounded by that of a sum of b independent geometric random variables:

$$\mathbb{E}\left\{T^2 \,|\, \boldsymbol{\Theta}(t_g)\right\} \leq b(1-\phi)/\phi^2 + b^2/\phi^2$$

Proof: (Lemma 1) The proof follows by squaring the queue update equations (1) and (15) and using a multi-slot drift analysis. See Appendix A for the full proof. □

Let $V \geq 0$ be a non-negative parameter that we shall use to affect proximity to the optimal solution (with a tradeoff in convergence times and average queue congestion, as shown below). For pure feasibility problems, we set $V = 0$. As in [1] [18] for the case of single-slot problems, our strategy in this variable slot scenario is, upon every renewal event, to take control actions that minimize the following "drift-plus-penalty" expression:

$$D(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1} x_0(t_g+\tau) \,|\, \boldsymbol{\Theta}(t_g)\right\} \qquad (20)$$

where T is the random time until the next renewal event.

Given the queue backlogs $\boldsymbol{\Theta}(t_g)$ at the start of the renewal time t_g, the expression (20) represents a sum of random drift and penalty terms (which depend on control actions) over the course of a renewal interval. Hence, controlling the system to minimize this sum amounts to solving a *weighted stochastic shortest path problem* over the renewal interval (see [2] for a treatment of the theory of stochastic shortest path problems). This generalizes the well known max-weight policies of [1] [18] [19] [24]. Indeed, in [1] [18] [19] [24] there is no $z(t)$ state and so "renewals" occur every slot and the shortest path problem reduces to a simple greedy control action that minimizes a weighted drift-plus-penalty term over one slot. In this generalization, the queue backlogs still act as weights, but the solution of the stochastic shortest path problem is not greedy and requires consideration of how an action at time t affects the Markov state $z(t)$ in future slots $t \geq t_g$.

B. Feasibility Problems ($V = 0$)

Suppose that we have a pure feasibility problem. Define $V = 0$, and define renewals according to any valid definition (such as the type-1 definition). Suppose that there are constants $C \geq 0$ and $\delta \geq 0$ such that upon every renewal time t_g, we observe the queue backlogs $\boldsymbol{\Theta}(t_g)$ and make control decisions

[5] Note that proper notation for the drift should be $\Delta_T(\boldsymbol{\Theta}(t_g), t_g)$, as the drift may result from a non-stationary policy and hence can depend on the starting time t_g, although we use the simpler notation $\Delta_T(\boldsymbol{\Theta}(t_g))$ as a formal representation of the right hand side of (16).

over the course of the renewal interval that satisfy:

$$\begin{aligned} D(\boldsymbol{\Theta}(t_g)) \quad &\leq \quad D^{ssp}(\boldsymbol{\Theta}(t_g)) + C + \delta \sum_{n\in\mathcal{N}} Q_n(t_g) \\ &\qquad + \delta \sum_{m\in\mathcal{M}} Y_m(t_g) \end{aligned} \tag{21}$$

where $D^{ssp}(\boldsymbol{\Theta}(t_g))$ denotes the value of the expression (20) under the optimal solution to the stochastic shortest path problem (which would take place over a random renewal duration T^{ssp} that depends on the random events that would occur under this solution), and where $D(\boldsymbol{\Theta}(t_g))$ denotes the corresponding value of (20) under the actual control actions taken (with a random renewal duration T that depends on the actual random events that occur). Note that if the exact stochastic shortest path solution is implemented every renewal interval, we have $C = 0$ and $\delta = 0$.

Theorem 1: (Performance for Feasibility Problems) Suppose Assumption 1 is satisfied for a given $\epsilon > 0$. If there are constants $C \geq 0$ and $\delta \geq 0$ such that (21) is satisfied for every renewal interval, and if δ is small enough so that:

$$\epsilon \mathbb{E}\left\{T^*\right\} > \delta \tag{22}$$

where $\mathbb{E}\left\{T^*\right\}$ is the expected renewal duration under the policy $I^*(t)$ from Assumption 1, then all queues $Q_n(t)$ and $Y_m(t)$ are *strongly stable*, in the sense that $\overline{Q}_n < \infty$ and $\overline{Y}_m < \infty$ for all $n \in \mathcal{N}$ and $m \in \mathcal{M}$. Consequently, all time average feasibility constraints (3)-(4) are satisfied. Furthermore, define t_g as the timeslot of the gth renewal (for $g \in \{0, 1, 2, \ldots\}$ and $t_0 = 0$). Then the time average expectation of queue backlogs over the first G renewal intervals satisfies (for any integer $G > 0$):

$$\begin{aligned} \frac{1}{G}\sum_{g=0}^{G-1}\left[\sum_{n\in\mathcal{N}}\mathbb{E}\left\{Q_n(t_g)\right\} + \sum_{m\in\mathcal{M}}\mathbb{E}\left\{Y_m(t_g)\right\}\right] \\ \leq \frac{B+C}{\epsilon\mathbb{E}\left\{T^*\right\} - \delta} \end{aligned} \tag{23}$$

where we recall that for type-1 renewals $\mathbb{E}\left\{T^*\right\}$ satisfies $1 \leq \mathbb{E}\left\{T^*\right\} \leq 1/\phi$, and for type-2 renewals $\mathbb{E}\left\{T^*\right\} = 1/\phi$.

Proof: We first prove (23). From (17) and (21) we have for any renewal time t_g:

$$\begin{aligned} \Delta_T(\boldsymbol{\Theta}(t_g)) \leq B + C + D^{ssp}(\boldsymbol{\Theta}(t_g)) + \delta \sum_{n\in\mathcal{N}} Q_n(t_g) \\ + \delta \sum_{m\in\mathcal{M}} Y_m(t_g) \end{aligned}$$

However, by definition, the stochastic shortest path policy yields a value of $D^{ssp}(\boldsymbol{\Theta}(t_g))$ that is less than or equal to the corresponding value under any other policy, and hence:

$$D^{ssp}(\boldsymbol{\Theta}(t_g)) \leq D^*(\boldsymbol{\Theta}(t_g))$$

where $D^*(\boldsymbol{\Theta}(t_g))$ represents the value under any other policy that could be implemented over a renewal interval that starts with queue backlogs $\boldsymbol{\Theta}(t_g)$ (where the renewal interval for this other policy may have a different duration than that of the stochastic shortest path policy). Thus:

$$\begin{aligned} \Delta_T(\boldsymbol{\Theta}(t_g)) \quad &\leq \quad B + C + D^*(\boldsymbol{\Theta}(t_g)) \\ &\qquad + \delta \sum_{n\in\mathcal{N}} Q_n(t_g) + \delta \sum_{m\in\mathcal{M}} Y_m(t_g) \\ &= B + C \\ &\quad - \sum_{n\in\mathcal{N}} Q_n(t_g)\mathbb{E}\left\{-\delta + \sum_{\tau=0}^{T^*-1} d_n^*(t_g+\tau)\right\} \\ &\quad - \sum_{m\in\mathcal{M}} Y_m(t_g)\mathbb{E}\left\{-\delta + T^* x_m^{av} - \sum_{\tau=0}^{T^*-1} x_m^*(t_g+\tau)\right\} \end{aligned} \tag{24}$$

where the final equality holds by definition of $D^*(\boldsymbol{\Theta}(t_g))$ in (18). Now consider the (z, Ω)-only policy $I^*(t)$ of Assumption 1, which satisfies inequalities (10)-(11) for some value $\epsilon > 0$, and which also makes decisions independent of $\boldsymbol{\Theta}(t_g)$. Plugging (10)-(11) into (24) yields:

$$\begin{aligned} \Delta_T(\boldsymbol{\Theta}(t_g)) \quad &\leq \quad B + C - (\epsilon\mathbb{E}\left\{T^*\right\} - \delta) \sum_{n\in\mathcal{N}} Q_n(t_g) \\ &\qquad - (\epsilon\mathbb{E}\left\{T^*\right\} - \delta) \sum_{m\in\mathcal{M}} Y_m(t_g) \end{aligned} \tag{25}$$

The above holds for any $g \in \{0, 1, 2, \ldots\}$. Taking expectations of both sides of (25) and using the definition of $\Delta(\boldsymbol{\Theta}(t_g))$ given in (16) yields:

$$\begin{aligned} \mathbb{E}\left\{L(\boldsymbol{\Theta}(t_{g+1}))\right\} - \mathbb{E}\left\{L(\boldsymbol{\Theta}(t_g))\right\} \quad &\leq \quad B + C \\ -(\epsilon\mathbb{E}\left\{T^*\right\} - \delta)\left[\sum_{n\in\mathcal{N}}\mathbb{E}\left\{Q_n(t_g)\right\} + \sum_{m\in\mathcal{M}}\mathbb{E}\left\{Y_m(t_g)\right\}\right] \end{aligned}$$

Summing the above inequality over $g \in \{0, 1, \ldots, G-1\}$ and using the fact that all queues are initially empty (so that $L(\boldsymbol{\Theta}(0)) = 0$), and dividing by G yields:

$$\begin{aligned} \frac{\mathbb{E}\left\{L(\boldsymbol{\Theta}(t_G))\right\}}{G} &\leq (B+C) \\ -\frac{\epsilon\mathbb{E}\left\{T^*\right\} - \delta}{G}&\sum_{g=0}^{G-1}\left[\sum_{n\in\mathcal{N}}\mathbb{E}\left\{Q_n(t_g)\right\} + \sum_{m\in\mathcal{M}}\mathbb{E}\left\{Y_m(t_g)\right\}\right] \end{aligned}$$

Rearranging terms and using non-negativity of $L(\boldsymbol{\Theta}(t_G))$ together with the fact that $\epsilon\mathbb{E}\left\{T^*\right\} - \delta > 0$ (by the assumption in (22)) yields the result of (23).

The fact that the queues $Q_n(t)$ and $Y_m(t)$ are strongly stable follows as a simple consequence of (23) together with the facts that (i) queue backlog growth is deterministically bounded every slot, and (ii) first and second moments of renewal times are bounded. This is formally shown in Appendix B for completeness. □

Note that type-1 renewals are the best for feasibility problems, as they have the smallest renewal duration and thus have smaller values of the B constant (and hence smaller average sizes for queues $Q_n(t)$ and $Y_m(t)$). Indeed, the definition of B and the bound on $\mathbb{E}\left\{T^2 \mid \boldsymbol{\Theta}(t_g)\right\}$ given in Lemma 1 imply the following for type-1 renewals:

$$B \leq \frac{(2-\phi)\sigma^2}{2\phi^2} \tag{26}$$

This is compared to the type-3 renewals which consist of $b > 1$ type-1 renewals and yield:

$$B \leq \frac{(b(1-\phi) + b^2)\sigma^2}{2\phi^2} \tag{27}$$

However, type-3 renewals are useful in cases when the shortest path problem is solved using online approximation techniques based on forward simulation, such as the Q-learning algorithms in [2], which require a longer time to converge to a solution that satisfies the approximation bound (21). Using $b > 1$ in this way can be viewed as a kind of 2-timescale approach, where b is proportional to the timescale required for accuracy of the stochastic shortest path approximation. One difficulty with this approach is that the value b required to obtain an accurate approximation may be geometric in K, which then creates an exponential bound on B in (27). An alternative (single-timescale) approximation technique that does not require $b > 1$ and that preserves the polynomial bound of (26) associated with type-1 renewals (i.e., $b = 1$) is provided in Section IV.

C. Optimization Problems ($V > 0$)

Consider now the optimization problem (2)-(4), so that we desire to minimize the time average of the penalty $x_0(t)$, and the V parameter in the stochastic shortest path problem (20) is positive. Further suppose that our renewals are defined as type-2 renewals, so that renewal events are only at forced renewal times, which are i.i.d. Bernoulli with probability ϕ. Suppose that there are constants $C \geq 0$ and $\delta \geq 0$ such that on every renewal interval we observe the queue states $\boldsymbol{\Theta}(t_g)$ and take actions that satisfy the following approximation:

$$\begin{aligned} D(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1} x_0(t_g+\tau) \mid \boldsymbol{\Theta}(t_g)\right\} \leq \\ D^{ssp}(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1} x_0^{ssp}(t_g+\tau) \mid \boldsymbol{\Theta}(t_g)\right\} \\ +C + \delta\sum_{n\in\mathcal{N}} Q_n(t_g) + \delta\sum_{m\in\mathcal{M}} Y_m(t_g) + V\delta \end{aligned} \tag{28}$$

where $x_0(t)$ represents the penalty that is incurred by the implemented policy, $x_0^{ssp}(t)$ is the penalty that would be incurred under the stochastic shortest path solution to (20), and T is the renewal frame size (which is unaffected by control decisions for type-2 renewals and satisfies $\mathbb{E}\{T\} = 1/\phi$). Note that if the exact stochastic shortest path solution to (20) is used every renewal interval, we have $C = \delta = 0$.

Theorem 2: (Performance for Optimization Problems) Suppose we use type-2 renewals (where all renewals are forced renewals and occur i.i.d. with probability ϕ), and suppose Assumptions 1 and 2 hold for a given $\epsilon > 0$. Fix a parameter $V > 0$. If there are constants $C \geq 0$ and $\delta \geq 0$ such that (28) is satisfied for every renewal interval, and if δ is small enough so that $\epsilon > \phi\delta$, then $\overline{Q}_n < \infty$ and $\overline{Y}_m < \infty$ for all $n \in \mathcal{N}$ and $m \in \mathcal{M}$ (and consequently feasibility constraints (3)-(4) are satisfied). Furthermore, for renewal times t_g (for $g \in \{0, 1, 2, \ldots\}$) and for any positive integer G we have:

$$\begin{aligned} \frac{1}{G}\sum_{g=0}^{G-1}\left[\sum_{n\in\mathcal{N}}\mathbb{E}\{Q_n(t_g)\} + \sum_{m\in\mathcal{M}}\mathbb{E}\{Y_m(t_g)\}\right] \leq \\ \frac{(B+C)\phi + V(\phi\delta + x_0^{max} - x_0^{min})}{\epsilon - \phi\delta} \end{aligned} \tag{29}$$

Finally, the time average penalty satisfies:

$$\begin{aligned} \limsup_{t\to\infty} \tfrac{1}{t}\textstyle\sum_{\tau=0}^{t-1}\mathbb{E}\{x_0(\tau)\} \leq x_0^{opt} + \\ \tfrac{(B+C)\phi}{V} + \phi\delta[1 + (x_0^{max} - x_0^{opt})/\epsilon] \end{aligned} \tag{30}$$

and the right-hand side is also a bound on the average penalty over G renewal intervals divided by the average duration of G renewal intervals, for any positive G (see equation (81) in Appendix C).

Note from (30) and (29) that the time average of $x_0(t)$ can be made arbitrarily close to (or below) $x_0^{opt} + \phi\delta[1 + (x_0^{max} - x_0^{opt})/\epsilon]$ as V is increased, with a tradeoff in average queue size that is linear in V. The value δ determines how close this performance is to the optimal value x_0^{opt}. In the case $\delta = 0$ (which holds, for example, if our approximation to the stochastic shortest path problem differs from the optimal solution only by a constant C that is independent of queue length), then the V parameter affects a $[O(1/V); O(V)]$ performance-delay tradeoff, as in [1], so that distance to optimality is $O(1/V)$ and hence can be made arbitrarily small, at the expense of an increase in the average backlog of the queues that is linear in V. This average backlog of the queues $\boldsymbol{Q}(t)$ directly affects their average delay (via Little's Theorem), while the average backlog of the virtual queues $Y_m(t)$ affects the average *convergence time* required to achieve the performance guarantees (see also, for example, [25]).

Proof: See Appendix C. □

Only type-2 renewals are considered in Theorem 2 because $\mathbb{E}\{T\} = \mathbb{E}\{T^*\} = 1/\phi$ for type-2 renewal intervals, a property which is needed for (30).

IV. Solving the Stochastic Shortest Path Problem

Consider now the stochastic shortest path problem given by expression (20). Here we describe its solution, using either a type-1 or type-2 renewal definition (so that renewals always occur at forced renewal events, and have mean duration at most $1/\phi$). Without loss of generality, assume we start at time 0 and have (possibly non-zero) backlogs $\boldsymbol{\Theta} = \boldsymbol{\Theta}(0)$. Let T be the renewal interval size. For every step $\tau \in \{0, \ldots, T-1\}$, define $c_{\boldsymbol{\Theta}}(I(\tau), \Omega(\tau), z(\tau))$ as the incurred cost assuming that the queue state at the beginning of the renewal is $\boldsymbol{\Theta}(0)$:

$$\begin{aligned} c_{\boldsymbol{\Theta}}(I(\tau), \Omega(\tau), z(\tau)) \triangleq -\sum_{n\in\mathcal{N}} Q_n(0)\hat{d}_n(I(\tau), \Omega(\tau), z(\tau)) \\ -\sum_{m\in\mathcal{M}} Y_m(0)[x_m^{av} - \hat{x}_m(I(\tau), \Omega(\tau), z(\tau))] \\ +V\hat{x}_0(I(\tau), \Omega(\tau), z(\tau)) \end{aligned} \tag{31}$$

Let $I^{ssp}(\tau)$ denote the optimal control action on slot τ for solving the stochastic shortest path problem, given that the controller first observes $\Omega(\tau)$ and $z(\tau)$. Define $\mathcal{Z}_r \triangleq \mathcal{Z} \cup \{r\}$, where we have added a new state "r" to represent the renewal state, which is the termination state of the stochastic shortest path problem. Appropriately adjust the probability transition matrix $P = (P_{zy}(I,\Omega))$ to account for this new state [26] [2]. For example, for type-1 renewals, all transition probabilities are the same as before with the exception that transitions to state 0 are replaced with transitions to r. Define $\boldsymbol{J} = (J_z)|_{z\in\mathcal{Z}_r}$ as a vector of optimal costs, where J_z is the minimum expected sum cost to the renewal state given that we start in state z, and $J_r = 0$. By basic dynamic programming theory [26] [2], the optimal control action on each slot τ (given $\Omega(\tau)$ and $z(\tau)$) is:

$$I(\tau) = \arg\min_{I\in\mathcal{I}_{\Omega(\tau),z(\tau)}} [c_{\boldsymbol{\Theta}}(I,\Omega(\tau),z(\tau))+ \sum_{y\in\mathcal{Z}_r} P_{z(\tau),y}(I,\Omega(\tau))J_y] \quad (32)$$

This policy is easily implemented provided that the J_z values are known. It is well known that the $\boldsymbol{J}$ vector satisfies the following vector dynamic programming equation:[6]

$$\boldsymbol{J} = \mathbb{E}\left\{\min_{I\in\mathcal{I}_{\Omega(\tau),\boldsymbol{z}}} [\boldsymbol{c}_{\boldsymbol{\Theta}}(I,\Omega(\tau)) + P(I,\Omega(\tau))\boldsymbol{J}]\right\} \qquad (33)$$

where we have used an entry-wise min (possibly with different I vectors being used for minimizing each entry $z\in\mathcal{Z}$). Thus, the notation $I\in\mathcal{I}_{\Omega(t),\boldsymbol{z}}$ emphasizes that for a given $z\in\mathcal{Z}$, the control action I is chosen from the set $\mathcal{I}_{\Omega(t),z}$. Further, $\boldsymbol{c}_{\boldsymbol{\Theta}}(I,\Omega(t))$ is defined as a vector with entries $c_{z,\boldsymbol{\Theta}}(I,\Omega(t)) = c_{\boldsymbol{\Theta}}(I,\Omega(t),z)$, and $P(I,\Omega(t)) = (P_{zy}(I,\Omega))$ is the probability transition matrix under $\Omega(t)$ and control action I. The expectation in (33) is over the distribution of the i.i.d. process $\Omega(t)$. Because $\Omega(t)$ has the structure $\Omega(t) = [\omega(t);\phi(t)]$ where $\omega(t)$ is the random outcome for slot t and $\phi(t)$ is an independent Bernoulli process that has forced renewals with probability ϕ, we can re-write the above vector equation as:

$$\boldsymbol{J} = \phi\mathbb{E}\left\{\min_{I\in\mathcal{I}_{[\omega(t),1],\boldsymbol{z}}} \boldsymbol{c}_{\boldsymbol{\Theta}}^{(1)}(I,\omega(t))\right\} + (1-\phi)\mathbb{E}\left\{\min_{I\in\mathcal{I}_{[\omega(t),0],\boldsymbol{z}}} \left[\boldsymbol{c}_{\boldsymbol{\Theta}}^{(0)}(I,\omega(t)) + P^{(0)}(I,\omega(t))\boldsymbol{J}\right]\right\} \quad (34)$$

where:

$$\begin{aligned} \boldsymbol{c}_{\boldsymbol{\Theta}}^{(1)}(I,\omega(t)) &\triangleq \boldsymbol{c}_{\boldsymbol{\Theta}}(I,[\omega(t),1]) \\ \boldsymbol{c}_{\boldsymbol{\Theta}}^{(0)}(I,\omega(t)) &\triangleq \boldsymbol{c}_{\boldsymbol{\Theta}}(I,[\omega(t),0]) \\ P^{(0)}(I,\omega(t)) &\triangleq P(I(t),[\omega(t),0]) \end{aligned}$$

We assume that the probability transition matrix $P^{(0)}(I,\omega(t))$ is known (recall that this is indeed a known $0/1$ matrix in the case of the system with delay-constrained and delay-unconstrained users of Section II-C). We next show how to compute an approximation of $\boldsymbol{J}$ based on random samples of $\omega(t)$ and using a classic Robbins-Monro iteration.

A. Estimation Through Random i.i.d. Samples

Suppose we have an infinite sequence of random variables arranged in batches with batch size L, with ω_{bi} denoting the ith sample of batch b. All random variables are i.i.d. with probability distribution the same as $\omega(t)$, and all are independent of the queue state $\boldsymbol{\Theta}$ that is used for this stochastic shortest path problem. Consider the following two mappings Ψ and $\tilde{\Psi}$ from a $\boldsymbol{J}$ vector to another $\boldsymbol{J}$ vector, where the second is implemented with respect to a particular batch b:

$$\Psi\boldsymbol{J}\triangleq\phi\mathbb{E}\left\{\min_{I\in\mathcal{I}_{[\omega(t),1],\boldsymbol{z}}} \boldsymbol{c}_{\boldsymbol{\Theta}}^{(1)}(I,\omega(t))\right\} + (1-\phi)\mathbb{E}\left\{\min_{I\in\mathcal{I}_{[\omega(t),0],\boldsymbol{z}}} \left[\boldsymbol{c}_{\boldsymbol{\Theta}}^{(0)}(I,\omega(t)) + P^{(0)}(I,\omega(t))\boldsymbol{J}\right]\right\} \quad (35)$$

$$\tilde{\Psi}\boldsymbol{J}\triangleq\phi\frac{1}{L}\sum_{i=1}^{L}\min_{I\in\mathcal{I}_{[\omega_{bi},1],\boldsymbol{z}}} \boldsymbol{c}_{\boldsymbol{\Theta}}^{(1)}(I,\omega_{bi}) + (1-\phi)\frac{1}{L}\sum_{i=1}^{L}\min_{I\in\mathcal{I}_{[\omega_{bi},0],\boldsymbol{z}}} \left[\boldsymbol{c}_{\boldsymbol{\Theta}}^{(0)}(I,\omega_{bi}) + P^{(0)}(I,\omega_{bi})\boldsymbol{J}\right] \quad (36)$$

where the min is entrywise over each vector entry. The expectation in (35) is implicitly conditioned on a given $\boldsymbol{\Theta}$ vector, and is with respect to the random $\boldsymbol{\omega}(t)$ event that is independent of $\boldsymbol{\Theta}$. The mapping Ψ cannot be implemented without knowledge of the distribution of $\omega(t)$ (so that the expectation can be computed), whereas the mapping $\tilde{\Psi}$ can be implemented as a "simulation" over the L random samples ω_{bi} (assuming such samples can be generated or obtained). Note however that the expected value of $\tilde{\Psi}\boldsymbol{J}$ is exactly equal to $\Psi\boldsymbol{J}$. Thus, given an initial vector $\boldsymbol{J}_b$ for use for step b (with some initial guess for $\boldsymbol{J}_0$, such as $\boldsymbol{J}_0 = \boldsymbol{0}$), we can write $\tilde{\Psi}\boldsymbol{J}_b = \Psi\boldsymbol{J}_b + \boldsymbol{\eta}_b$, where $\boldsymbol{\eta}_b$ is a zero-mean vector random variable. Specifically, the vector $\boldsymbol{\eta}_b$ satisfies:

$$\mathbb{E}\left\{\boldsymbol{\eta}_b \mid \boldsymbol{J}_b\right\} = \boldsymbol{0}$$

Thus, while the vector $\boldsymbol{\eta}_b$ is *not* independent of $\boldsymbol{J}_b$, each entry is *uncorrelated* with any deterministic function of $\boldsymbol{J}_b$. That is, for each entry i and any deterministic function $f(\cdot)$ we have via iterated expectations:

$$\mathbb{E}\left\{\eta_b[i]f(\boldsymbol{J}_b)\right\} = \mathbb{E}\left\{f(\boldsymbol{J}_b)\mathbb{E}\left\{\eta_b[i] \mid \boldsymbol{J}_b\right\}\right\} = 0 \qquad (37)$$

For $b\in\{0,1,2,\ldots\}$ we have the iteration:

$$\boldsymbol{J}_{b+1} = \frac{1}{b+1}\tilde{\Psi}\boldsymbol{J}_b + \frac{b}{b+1}\boldsymbol{J}_b \qquad (38)$$

This iteration is a classic *Robbins-Monro* stochastic approximation algorithm. It can be shown that the $\boldsymbol{J}$ vector remains deterministically bounded for all b (see Lemma 2 below), and that Ψ and $\tilde{\Psi}$ satisfy the requirements of Proposition 4.6 in Section 4.3.4 of [2]. Thus the above iteration is in the standard form for stochastic approximation theory, and ensures that:

$$\lim_{b\rightarrow\infty}\boldsymbol{J}_b = \boldsymbol{J}^* \quad \text{with prob. } 1$$

where $\boldsymbol{J}^*$ is the cost vector associated with the optimal stochastic shortest path problem, that is, it is the solution to (34) and thus satisfies $\boldsymbol{J}^* = \Psi\boldsymbol{J}^*$. This holds for any batch size L (including the simplest case $L = 1$), although taking larger batches may improve overall convergence as the variance of the per-batch estimation is lower.

However, because our estimates do not need to converge to the exact value of $\boldsymbol{J}^*$, we modify the iteration (38) as follows:

$$\boldsymbol{J}_{b+1} = \gamma\tilde{\Psi}\boldsymbol{J}_b + (1-\gamma)\boldsymbol{J}_b \qquad (39)$$

[6] One can also derive (33) by defining a value function $H(z,\Omega)$, writing the Bellman equation in terms of $H(z(t{+}1),\Omega(t{+}1))$, taking an expectation with respect to the i.i.d. $\Omega(t)$, $\Omega(t+1)$, and defining $J(z)\triangleq\mathbb{E}_{\Omega(t)}\{H(z,\Omega(t))\}$.

where γ is a value such that $0 < \gamma < 1$, chosen to be suitably small to provide an accurate approximation, as specified below. We first define a norm for random vectors.

Definition 2: Let $\boldsymbol{X} = (X_j)$ be a random vector. We define the *entrywise root expected square norm* (or e-norm) $||\boldsymbol{X}||_e$ and its *deterministic* version $||\boldsymbol{X}||_d$ as follows:

$$\begin{aligned} ||\boldsymbol{X}||_e &\triangleq \sqrt{\max_j \mathbb{E}\left\{X_j^2\right\}} \\ ||\boldsymbol{X}||_d &\triangleq \max_j |X_j| \end{aligned}$$

It is easy to verify that for any random vector $\boldsymbol{X}$ we have $||\boldsymbol{X}||_e \geq 0$, $||\boldsymbol{X}||_e = ||-\boldsymbol{X}||_e$, $||\alpha\boldsymbol{X}||_e = \alpha||\boldsymbol{X}||_e$ for any non-negative scalar α, and the triangle inequality $||\boldsymbol{X}+\boldsymbol{Y}||_e \leq ||\boldsymbol{X}||_e + ||\boldsymbol{Y}||_e$ holds for all random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ with the same dimension. If $\boldsymbol{X}$ is a deterministic constant vector, then $||\boldsymbol{X}||_d = ||\boldsymbol{X}||_e$.

Now consider the iteration (39). We want to show that the noise vectors $\{\boldsymbol{\eta}_b\}_{b=0}^{\infty}$ are bounded for all b. To this end, let c_{max} denote the maximum absolute value of any entry of the $\boldsymbol{c}_{\boldsymbol{\Theta}}^{(1)}(I, \omega(t))$ and $\boldsymbol{c}_{\boldsymbol{\Theta}}^{(0)}(I, \omega(t))$ vectors, considering all possible $\omega(t)$ and I. Note that this value is finite due to the finiteness of the penalty functions. The size of c_{max} grows linearly in V and in the size of queue backlogs $\boldsymbol{\Theta}$.

Lemma 2: Define $J_{max} \triangleq c_{max}/\phi$. If $||\boldsymbol{J}_0||_d \leq J_{max}$, then:

(a) For all $b \in \{0, 1, 2, \ldots\}$ we have:

$$||\boldsymbol{J}_b||_d \leq J_{max}$$

(b) There are finite constants η_{min} and η_{max} such that $\eta_{min} \leq \eta_b[i] \leq \eta_{max}$ for all iterations b and all entries i. Further, for all b, if batches of size $L \geq 1$ are used, then:

$$||\boldsymbol{\eta}_b||_e^2 \leq \frac{|\eta_{min}\eta_{max}|}{L} \leq \frac{4(c_{max} + (1-\phi)J_{max})^2}{L}$$

Proof: See Appendix E. □

Lemma 3: Let $\boldsymbol{J}_b$ be the bth iteration of (39), starting with some initial vector $\boldsymbol{J}_0$ with $||\boldsymbol{J}_0||_e \leq J_{max}$. Assume that for all b we have $||\boldsymbol{\eta}_b||_e^2 \leq \sigma^2$ for some finite constant σ^2 (as in Lemma 2 with $\sigma^2 = |\eta_{min}\eta_{max}|/L$). Let $\boldsymbol{J}^*$ be the optimal solution to (34), satisfying $\Psi\boldsymbol{J}^* = \boldsymbol{J}^*$. Then:

(a) Every one-step iteration satisfies (for integers $b \geq 0$):

$$||\boldsymbol{J}_{b+1} - \boldsymbol{J}^*||_e^2 \leq (1-\phi\gamma)^2||\boldsymbol{J}_b - \boldsymbol{J}^*||_e^2 + \gamma^2||\boldsymbol{\eta}_b||_e^2$$

(b) After b iterations we have:

$$\begin{aligned} ||\boldsymbol{J}_b - \boldsymbol{J}^*||_e^2 \leq{}& (1-\phi\gamma)^{2b}||\boldsymbol{J}_0 - \boldsymbol{J}^*||_e^2 \\ &+ \frac{\gamma\sigma^2(1-(1-\phi\gamma)^{2b})}{\phi(2-\phi\gamma)} \end{aligned}$$

(c) In the limit, we have:

$$\lim_{b\rightarrow\infty} ||\boldsymbol{J}_b - \boldsymbol{J}^*||_e^2 \leq \frac{\gamma\sigma^2}{\phi(2-\phi\gamma)}$$

Proof: See Appendix E. □

Part (c) shows that the limiting deviation from the desired $\boldsymbol{J}^*$ vector can be made as small as desired by choosing a suitably small value of γ (and a suitably large value of b). We now show that an implementation that chooses $I(t)$ over a frame according to (32), using the $\boldsymbol{J}_b$ estimate instead of the optimal $\boldsymbol{J}^*$ vector, results in an approximation to the stochastic shortest path problem that deviates by an amount that depends on $||\boldsymbol{J}_b - \boldsymbol{J}^*||_e$.

Lemma 4: Suppose we choose $I(t)$ according to (32) over the course of a frame, using a (possibly random) vector $\boldsymbol{J}$ rather than $\boldsymbol{J}^*$. Define $\tilde{\boldsymbol{J}}(\boldsymbol{J})$ as the vector of expected sum costs over the frame (given this implementation uses $\boldsymbol{J}$). Then:

$$||\tilde{\boldsymbol{J}}(\boldsymbol{J}) - \boldsymbol{J}^*||_e \leq \frac{2(1-\phi)||\boldsymbol{J} - \boldsymbol{J}^*||_e}{\phi}$$

Further, defining $\boldsymbol{1}$ as a vector of all 1 entries with size equal to the dimension of $\boldsymbol{J}$, we have:

$$\mathbb{E}\left\{\tilde{\boldsymbol{J}}(\boldsymbol{J})\right\} \leq \boldsymbol{J}^* + \boldsymbol{1}\frac{2(1-\phi)||\boldsymbol{J} - \boldsymbol{J}^*||_e}{\phi}$$

where the expectation above is with respect to the randomness of the $\boldsymbol{J}$ vector.

Proof: See Appendix E □

B. Choosing b and γ

Lemma 3 can be used to compute the C and δ constants in Theorems 1 and 2. For example, if we start the iterations with $\boldsymbol{J}_0 = \boldsymbol{0}$, and note that $||\boldsymbol{J}^*||_e \leq J_{max} = c_{max}/\phi$, we find from part (b) of Lemma 3 that the main error term decreases exponentially fast (in b), so that:

$$||\boldsymbol{J}_b - \boldsymbol{J}^*||_e \leq \sqrt{(1-\phi\gamma)^{2b}\frac{c_{max}^2}{\phi^2} + \frac{\gamma\sigma^2(1-(1-\phi\gamma)^{2b})}{\phi(2-\phi\gamma)}} \tag{40}$$

Now choose b so that:

$$(1-\phi\gamma)^{2b}\frac{c_{max}^2}{\phi^2} - \frac{\gamma\sigma^2(1-\phi\gamma)^{2b}}{\phi(2-\phi\gamma)} \leq \frac{\gamma\sigma^2}{\phi(2-\phi\gamma)}$$

which implies from (40) that:

$$||\boldsymbol{J}_b - \boldsymbol{J}^*||_e \leq \sqrt{\frac{2\gamma\sigma^2}{\phi(2-\phi\gamma)}} \tag{41}$$

This is equivalent to choosing the integer b as follows: If $c_{max}^2/\phi^2 \leq \gamma\sigma^2/(\phi(2-\phi\gamma))$ then choose $b = 0$. Else, choose b such that:

$$b \geq \frac{\log(\frac{c_{max}^2(2-\phi\gamma)}{\gamma\sigma^2\phi} - 1)}{2\log(1/(1-\phi\gamma))} \tag{42}$$

With this choice of b, combining (41) with Lemma 4 shows that the expected cost over the duration of the frame when we use the $\boldsymbol{J}_b$ vector (rather than the optimal $\boldsymbol{J}^*$ vector) differs by no more than a constant α, where:

$$\alpha \triangleq \frac{2(1-\phi)\sqrt{2\gamma\sigma^2}}{\phi\sqrt{\phi(2-\phi\gamma)}} \leq \frac{2(1-\phi)\sqrt{2\gamma\sigma^2}}{\phi\sqrt{\phi(2-\phi)}}$$

Now note from (31) that c_{max} grows at most linearly in V and $||\boldsymbol{\Theta}||_d$. Hence (by Lemma 2b and Lemma 3), η_{min}, η_{max}, and $\sqrt{\sigma^2}$ also grow at most linearly and so:

$$\sqrt{\sigma^2} \leq d_L \max[||\boldsymbol{\Theta}||_d, V]$$

for some proportionality constant d_L that depends on the batch size L and the maximum penalties. Thus, for a given desired $\delta > 0$, choosing γ to satisfy:

$$0 < \gamma \leq \min\left[1, \frac{\delta^2\phi^3(2-\phi)}{8d_L^2(1-\phi)^2}\right]$$

ensures that:

$$\alpha \leq \max[||\boldsymbol{\Theta}||_d, V]\delta$$

Thus, inequalities (21) and (28) can be satisfied for this δ and choosing $C = 0$. This can be seen by simply placing the approximation error on $Q_n(t_g)\delta$, $Y_m(t_g)\delta$, or $V\delta$, depending on which term has the largest value. The above bound shows that γ must be chosen quite small, which requires a large number of iterations b according to (42) for provably high accuracy of the algorithm. However, the resulting b is still polynomial in the system parameters and in $1/\delta$, demonstrating that we need only a polynomial number of samples for performance to be arbitrarily close to the optimal with polynomial delay, and this is independent of the size of the state space $\mathcal{Z}$. In practice, one may not require such a small value of γ or a large value of b. Further, rather than starting the iterations with $\boldsymbol{J}_0 = \mathbf{0}$, performance can be significantly improved if we initiate the iterations on the current frame using the end value of $\boldsymbol{J}_b$ from the previous frame. The intuition is that the queue backlogs do not change significantly over the course of the frame, and hence the previous calculations can be exploited.

C. Complexity Discussion

Note that computing $\tilde{\Psi}\boldsymbol{J}$ in (36) involves taking a minimum over $I \in \mathcal{I}_{[\omega_{bi},0],\boldsymbol{z}}$. For the delay-constrained wireless example, this involves selecting one of the $K+N$ queues to serve, an operation with complexity that is *linear* in $K+N$. However, the minimization must be done for every entry of the $\boldsymbol{J}$ vector, the size of which is equal to the cardinality of the set $\mathcal{Z}$ (which is geometric in the number of delay-constrained queues K but independent of the number of delay-unconstrained queues N). This illustrates that we can solve problems with a very large number of delay-unconstrained queues, provided that the number of delay-constrained queues is small.

D. Sampling From the Past and Delayed Queue Analysis

It remains to be seen how one can obtain the required i.i.d. samples without knowing the probability distribution for $\omega(t)$. One might consider an online computation that obtains the samples by stepping *forward* in time. However, this requires a longer renewal interval to amortize the cost of learning the new samples, and hence creates additional congestion in the actual (delay-unconstrained) queues and in the virtual queues. In this subsection, we describe a technique that uses *previous* samples of the $\omega(\tau)$ values. This method maintains smaller renewal intervals and hence smaller bounds on the queues we are stabilizing.

We first obtain a collection of W i.i.d. samples of $\omega(t)$. Consider a given renewal time t_g, and suppose that the time t_g is large enough so that we can obtain W samples according to the following procedure: Let $\omega_1 \triangleq \omega(t_g)$. If we have a type-2 definition of renewals, we define: $\omega_2 \triangleq \omega(t_g - 1)$, $\omega_3 \triangleq \omega(t_g - 2), \ldots, \omega_W \triangleq \omega(t_g - W + 1)$. Because $\omega(t)$ is i.i.d. over slots (and because our type-2 renewals are chosen completely randomly), it is easy to see that $\{\omega_1, \ldots, \omega_W\}$ form an i.i.d. sequence. If we have a type-1 definition of renewals, we must be more careful in obtaining our samples, as the renewal times are not random but depend on past control decisions, which are correlated with the samples themselves. Nevertheless, we can begin finding samples at the last *forced renewal event*, and sample backwards in time from that point.

A subtlety now arises: Even though the $\{\omega_1, \ldots, \omega_W\}$ sequence is i.i.d., these samples are *not* independent of the queue backlog $\boldsymbol{\Theta}(t_g)$ at the beginning of the renewal. This is because these values have influenced the queue states. This makes it challenging to directly implement a Robbins-Monro iteration. Indeed, the expectation in (35) can be viewed as a conditional expectation given a certain queue backlog at the beginning of the renewal interval, which is $\boldsymbol{\Theta}(t_g)$ for the gth renewal. This conditioning does not affect (35) when $\omega(t)$ is chosen independently of initial queue backlog, and so the random samples in the Robbins-Monro iteration (39) are also assumed to be chosen independent of the initial queue backlog, which is not the case if we sample from the past.

To avoid this difficulty and ensure the samples are both i.i.d. and independent of the queue states that form the weights in our stochastic shortest path problem, we use a *delayed queue analysis*. Let t_{start} denote the slot on which sample ω_W is taken, and let $\boldsymbol{\Theta}(t_{start})$ represent the queue backlogs at that time. It follows that the i.i.d. samples are also independent of $\boldsymbol{\Theta}(t_{start})$. Hence, the bounds derived for the iteration technique in the previous section can be applied when the iterates use $\boldsymbol{\Theta}(t_{start})$ as the backlog vector. Let $\boldsymbol{J}_{\boldsymbol{\Theta}(t_g)}$ denote the optimal solution to the problem (33) for a queue backlog $\boldsymbol{\Theta}(t_g)$ at the beginning of our renewal time t_g, and let $\boldsymbol{J}_{\boldsymbol{\Theta}(t_{start})}$ denote the corresponding optimal solution for a problem that starts with initial queue backlog $\boldsymbol{\Theta}(t_{start})$. Let F denote the number of slots between t_{start} and t_g (so that $t_{start} + F = t_g$). For type-2 renewals, $F = W - 1$. For type-1 renewals, $F = H + W - 1$, where H is a geometric random variable with mean $1/\phi$. Because there are only F slots between time t_{start} and t_g, and the maximum change in any queue on one slot is bounded, we want to claim that the expected difference between these vectors is bounded. This is justified by the next lemma, which bounds the deviation of the optimal costs associated with two general queue backlog vectors.

Let $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$ be two different queue backlog vectors, and let $\boldsymbol{J}_{\boldsymbol{\Theta}_1}$ and $\boldsymbol{J}_{\boldsymbol{\Theta}_2}$ represent the optimal frame costs corresponding to $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$, respectively. Define the constant β as follows:

$$\beta \triangleq \sup_{I,\Omega} ||\boldsymbol{c}_{\boldsymbol{\Theta}_1}(I, \Omega) - \boldsymbol{c}_{\boldsymbol{\Theta}_2}(I, \Omega)||_d \tag{43}$$

where $\boldsymbol{c}_{\boldsymbol{\Theta}}(I, \Omega)$ is the vector, indexed by z, with the zth entry given by (31) using backlog vector $\boldsymbol{\Theta}$. Note from (31) that β is independent of V (as the V term in (31) cancels out in the subtraction), and is proportional to the maximum penalty value times the maximum *difference* in any queue backlog entry in $\boldsymbol{\Theta}_1$ and its corresponding entry in $\boldsymbol{\Theta}_2$. Thus β is also independent of the actual size of the backlog vectors, and depends only on their *difference*.

Lemma 5: For the vectors $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$, and for the β value defined in (43), we have:

(a) The difference between $\boldsymbol{J}_{\boldsymbol{\Theta}_1}$ and $\boldsymbol{J}_{\boldsymbol{\Theta}_2}$ satisfies:

$$||\boldsymbol{J}_{\boldsymbol{\Theta}_1} - \boldsymbol{J}_{\boldsymbol{\Theta}_2}||_d \leq \frac{\beta}{\phi}$$

(b) Let $I_1(t)$ denote the policy decisions at time t under the policy that makes optimal decisions subject to queue backlogs $\boldsymbol{\Theta}_1$, and define $\boldsymbol{J}_{21}^{mis}$ as the expected sum cost over a frame of a *mismatched policy* that incurs costs according to backlog vector $\boldsymbol{\Theta}_2$ but makes decisions according to $I_1(t)$ (and hence has the same frame duration and decisions as the optimal policy for $\boldsymbol{\Theta}_1$). Then:

$$\boldsymbol{J}_{\boldsymbol{\Theta}_2} \leq \boldsymbol{J}_{21}^{mis} \leq \boldsymbol{J}_{\boldsymbol{\Theta}_1} + \boldsymbol{1}\frac{\beta}{\phi}$$

where $\boldsymbol{1}$ is a vector of all 1 values with the same dimension as $\boldsymbol{J}_{\boldsymbol{\Theta}_1}$.

Proof: See Appendix G. □

The above lemma shows that if we use the Robbins-Monro iterations on $\boldsymbol{\Theta}(t_{start})$, then we achieve a vector that is close to $\boldsymbol{J}_{\boldsymbol{\Theta}(t_{start})}$ (according to the bounds given in the previous section), which is bounded by a constant from $\boldsymbol{J}_{\boldsymbol{\Theta}(t_g)}$, where the constant does not depend on the size of V and depends only on the maximum difference between queue backlogs $\boldsymbol{\Theta}(t_{start})$ and $\boldsymbol{\Theta}(t_g)$. Similar reasoning shows that the implementation of (32) can use any queue backlogs $\boldsymbol{\Theta}$ that are close to the queue backlogs at the start of the frame, *including using queue backlogs $\boldsymbol{\Theta}(t)$ that are updated on each slot in the frame*. While such implementation leads to a larger theoretical bound, in practice it may improve performance by allowing a faster reaction to emerging queue backlogs $Q_n(t)$.

V. Optimizing Convex Functions of Time Averages

Here we describe how optimization of convex functions of time averages can be achieved using the same framework of the previous sections. Specifically, we extend our method of *auxiliary variables* and *flow state queues*, developed for stochastic network optimization in [21] [1], to this Markov modulated network context. Consider the same network model as described in Section II. Define $\boldsymbol{x}(t)$ as a vector of penalties for $m \in \{1, \ldots, M\}$:

$$\boldsymbol{x}(t) \triangleq (x_1(t), \ldots, x_M(t))$$

where $x_m(t) = \hat{x}_m(I(t), \Omega(t), S(t))$. For a given policy $I(t)$, define the following t-slot time average:

$$\overline{\boldsymbol{x}}(t) \triangleq \frac{1}{t}\sum_{\tau=0}^{t-1} \mathbb{E}\left\{\boldsymbol{x}(\tau)\right\}$$

Rather than considering the stochastic network optimization problem (2)-(4), we consider a more general objective. Let $f(\boldsymbol{x})$, $h_1(\boldsymbol{x})$, $h_2(\boldsymbol{x}), \ldots, h_L(\boldsymbol{x})$ be a collection of continuous and convex functions of $\boldsymbol{x} \in \mathcal{R}^M$ (for some positive integer L). Define $\mathcal{L} \triangleq \{1, \ldots, L\}$. The generalized problem is:

$$\begin{aligned} \text{Minimize:} \quad & \limsup_{t\rightarrow\infty} f(\overline{\boldsymbol{x}}(t)) \\ \text{Subject to:} \quad & \limsup_{t\rightarrow\infty} h_l(\overline{\boldsymbol{x}}(t)) \leq c_l \text{ for all } l \in \mathcal{L} \\ & \overline{Q}_n < \infty \text{ for all } n \in \mathcal{N} \end{aligned}$$

where c_l are arbitrary constants.

This general objective is similar to the objectives of [1] [20] [21] which treat networks without the Markov modulated $z(t)$ variable. For simplicity of exposition, let us assume in this paragraph that all limits are well defined, and use $\overline{\boldsymbol{x}}$ to represent $\lim_{t\rightarrow\infty} \overline{\boldsymbol{x}}(t)$. With this notation, we can re-write the problem as:

$$\begin{aligned} \text{Minimize:} \quad & f(\overline{\boldsymbol{x}}) && (44) \\ \text{Subject to:} \quad & h_l(\overline{\boldsymbol{x}}) \leq c_l \text{ for all } l \in \mathcal{L} && (45) \\ & \overline{Q}_n < \infty \text{ for all } n \in \mathcal{N} && (46) \end{aligned}$$

This problem can be transformed as follows: Define $\boldsymbol{\gamma}(t) \triangleq (\gamma_1(t), \ldots, \gamma_M(t))$ as a vector of *auxiliary variables* (one auxiliary variable $\gamma_m(t)$ for each penalty $m \in \mathcal{M}$). On each slot t, $\boldsymbol{\gamma}(t)$ can be chosen as any vector that satisfies:

$$x_m^{min} - \alpha \leq \gamma_m(t) \leq x_m^{max} + \alpha \quad \forall m \in \mathcal{M} \tag{47}$$

for some fixed value $\alpha \geq 0$ (where choosing $\alpha > 0$ is sometimes useful for allowing slackness conditions). It is easy to see that the above problem is equivalent to the following:

$$\begin{aligned} \text{Minimize:} \quad & \overline{f(\boldsymbol{\gamma}(t))} && (48) \\ \text{Subject to:} \quad & \overline{h_l(\boldsymbol{\gamma}(t))} \leq c_l \;\; \forall l \in \mathcal{L} && (49) \\ & \overline{\gamma}_m = \overline{x}_m \;\; \forall m \in \mathcal{M} && (50) \\ & \overline{Q}_n < \infty \;\; \forall n \in \mathcal{N} && (51) \end{aligned}$$

where the time averages are defined:

$$\begin{aligned} \overline{f(\boldsymbol{\gamma}(t))} &\triangleq \lim_{t\rightarrow\infty} \frac{1}{t}\sum_{\tau=0}^{t-1} \mathbb{E}\left\{f(\boldsymbol{\gamma}(\tau))\right\} \\ \overline{h_l(\boldsymbol{\gamma}(t))} &\triangleq \lim_{t\rightarrow\infty} \frac{1}{t}\sum_{\tau=0}^{t-1} \mathbb{E}\left\{h_l(\boldsymbol{\gamma}(\tau))\right\} \end{aligned}$$

This equivalence can be briefly explained as follows: Let $\boldsymbol{x}(t)$ be the penalty vector used over time t under the optimal policy for the problem (44)-(46), and let $\overline{\boldsymbol{x}}$ be the time average (assumed to exist for simplicity of the discussion in this paragraph). Then we can use the same policy, together with the auxiliary variable decisions $\boldsymbol{\gamma}(t) = \overline{\boldsymbol{x}}$ for all t, to achieve the same penalty value and achieve the desired constraints in the new problem (48)-(51). Therefore, the minimum value in the new problem is less than or equal to that of the problem (44)-(46). On the other hand, any policy that optimizes the new problem can be found, by Jensen's inequality for convex functions, to also satisfy the constraints of the original problem (44)-(46), and to have a minimum value that is greater than or equal to the optimal value of the problem (44)-(46).

The new problem (48)-(51) is of the form described in the previous sections of this paper, with the exception of the linear equality constraint (50). There are several ways of treating this equality constraint. In the case when the functions $f(\boldsymbol{\gamma})$ and $h_l(\boldsymbol{\gamma})$ are non-increasing in each entry γ_m, we can replace the constraint with $\overline{\gamma}_m \leq \overline{x}_m$ for each $m \in \mathcal{M}$ (as in [1]). If the non-increasing property does not hold, we can approximate each linear equality constraint with two linear inequality constraints constraints $\overline{\gamma}_m \leq \overline{x}_m + \hat{\epsilon}$ and $\overline{x}_m \leq \overline{\gamma}_m + \hat{\epsilon}$, for some small value $\hat{\epsilon} > 0$ that allows the required slackness assumptions (Assumption 1) of Section II-E to hold. A more elegant solution, which does not require an approximation, uses a *generalized virtual queue* (which can possibly take negative values) of the form:

$$W_m(t+1) = W_m(t) - \gamma_m(t) + x_m(t) \tag{52}$$

Stabilizing this virtual queue $W_m(t)$ can be done via a Lyapunov function in a manner similar to stabilizing the other queues $Q_n(t)$ and $Y_m(t)$, and ensures that the linear equality constraint is satisfied. This approach is used in [27] in the case without the Markov modulated $z(t)$ variable. In the next section we combine this approach with our framework of Markov modulated networks with variable length frames.

A. The Generalized Weighted Shortest Path Algorithm

We have queues $Q_n(t)$ for $n \in \mathcal{N}$ with dynamics given by (1). For each $l \in \mathcal{L}$, define queues $Y_l(t)$ to enforce the constraints of (49):

$$Y_l(t+1) = \max[Y_l(t) - c_l + h_l(\boldsymbol{\gamma}(t)), 0] \tag{53}$$

We use the queues $W_m(t)$ in (52) to enforce the constraints (50). Define $\boldsymbol{\Theta}(t) \triangleq [\boldsymbol{Y}(t); \boldsymbol{Q}(t); \boldsymbol{W}(t)]$ as the combined queue vector, and define the following Lyapunov function:

$$L_2(\boldsymbol{\Theta}(t)) \triangleq \frac{1}{2}\left[\sum_{n\in\mathcal{N}} Q_n(t)^2 + \sum_{m\in\mathcal{M}} Y_m(t)^2 + \sum_{l\in\mathcal{L}} W_l(t)^2\right]$$

Consider any definition of a renewal. Let t_g be the start of a renewal time, and let T be its duration. Define:

$$\Delta_T(\boldsymbol{\Theta}(t_g)) \triangleq \mathbb{E}\left\{L_2(\boldsymbol{\Theta}(t_g+T)) - L_2(\boldsymbol{\Theta}(t_g)) \mid \boldsymbol{\Theta}(t_g)\right\}$$

Lemma 6: If t_g is a renewal time and T is the duration until the next renewal, we have:

$$\Delta_T(\boldsymbol{\Theta}(t_g)) \leq B_2 + D_2(\boldsymbol{\Theta}(t_g)) \tag{54}$$

where B_2 is a positive constant that satisfies for all t, $\boldsymbol{\Theta}(t_g)$:

$$\begin{aligned} B_2 \quad \geq \quad & \frac{\mathbb{E}\left\{T^2 \mid \boldsymbol{\Theta}(t_g)\right\}}{2}\left[\sum_{n\in\mathcal{N}}[\mu_n(t)^2 + R_n(t)^2] + \right. \\ & \left. \sum_{m\in\mathcal{M}}(\gamma_m(t) - x_m(t))^2 + \sum_{l\in\mathcal{L}}(h_l(\boldsymbol{\gamma}(t)) - c_l)^2\right] \end{aligned}$$

and $D_2(\boldsymbol{\Theta}(t_g))$ is defined:

$$\begin{aligned} D_2(\boldsymbol{\Theta}(t_g)) \triangleq & -\sum_{n\in\mathcal{N}} Q_n(t_g)\mathbb{E}\left\{\sum_{\tau=0}^{T-1} d_n(t_g+\tau) \mid \boldsymbol{\Theta}(t_g)\right\} \\ & -\sum_{m\in\mathcal{M}} W_m(t_g)\mathbb{E}\left\{\sum_{\tau=0}^{T-1}[\gamma_m(t_g+\tau) - x_m(t_g+\tau)] \mid \boldsymbol{\Theta}(t_g)\right\} \\ & -\sum_{l\in\mathcal{L}} Y_l(t_g)\mathbb{E}\left\{\sum_{\tau=0}^{T-1}[c_l - h_l(\boldsymbol{\gamma}(t_g+\tau))] \mid \boldsymbol{\Theta}(t_g)\right\} \end{aligned} \tag{55}$$

As before, our goal is to control the system over frames defined by renewal intervals starting at time t_g (and having duration T) by choosing $I(t)$ and $\gamma_m(t)$ subject to (47) to minimize the following drift-plus-penalty expression:

$$D_2(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1} f(\boldsymbol{\gamma}(t_g+\tau)) \mid \boldsymbol{\Theta}(t_g)\right\} \tag{56}$$

The Generalized Weighted Shortest Path Algorithm: Fix control parameters $V \geq 0$ and $\alpha \geq 0$. At time t_g, which is the beginning of the gth renewal interval (where $t_0 = 0$), observe queue values $\boldsymbol{\Theta}(t_g)$ and perform the following:

1) Compute $\boldsymbol{\gamma}_g = (\gamma_{1,g}, \ldots, \gamma_{M,g})$ that solves:

$$\begin{aligned} \text{Minimize:} \quad & Vf(\boldsymbol{\gamma}) - \textstyle\sum_{m\in\mathcal{M}} W_m(t_g)\gamma_m \\ & + \textstyle\sum_{l\in\mathcal{L}} Y_l(t_g)h_l(\boldsymbol{\gamma}) \\ \text{Subject to:} \quad & x_m^{min} - \alpha \leq \gamma_m \leq x_m^{max} + \alpha \ \forall m \in \mathcal{M} \end{aligned}$$

2) Choose $\boldsymbol{\gamma}(t_g+\tau) = \boldsymbol{\gamma}_g$ for all $\tau \in \{0, \ldots, T_g - 1\}$. That is, use the fixed value $\boldsymbol{\gamma}_g$ for the full duration of the gth renewal interval.
3) Choose control actions $I(t_g+\tau) \in \mathcal{I}_{\Omega(t_g+\tau), z(t_g+\tau)}$ over the renewal interval according to the optimal actions that solve the weighted stochastic shortest path problem (56) with $\boldsymbol{\gamma}(t_g+\tau) = \boldsymbol{\gamma}_g$. For each timestep $t_g+\tau$, update the actual queues $Q_n(t_g+\tau)$ according to (1) and update the virtual queues $Y_l(t_g+\tau)$ and $W_m(t_g+\tau)$ according to (53) and (52).

Note that the optimal solution to the weighted stochastic shortest path problem (56) has constant auxiliary variables $\boldsymbol{\gamma}(t_g+\tau) = \boldsymbol{\gamma}_g$ over the course of the renewal interval, where $\boldsymbol{\gamma}_g$ is computed in step 1. If the functions $f(\boldsymbol{\gamma})$ and $h_l(\boldsymbol{\gamma})$ are separable in each entry γ_m, the computation of step 1 reduces to finding, for each $m \in \mathcal{M}$, the minimum of a single-variable convex function over a closed interval.

B. Structure of the $f(\cdot)$ and $h_l(\cdot)$ functions

We assume the functions $f(\boldsymbol{\gamma})$ and $h_l(\boldsymbol{\gamma})$ (for $l \in \mathcal{L}$) are convex over the set of all $\boldsymbol{\gamma}$ vectors that satisfy (47). We further require the following mild assumptions.

- There are finite bounds f_{min} and f_{max} such that:
$$f_{min} \leq f(\boldsymbol{\gamma}) \leq f_{max} \text{ whenever } \boldsymbol{\gamma} \text{ satisfies (47)}$$
- Functions $f(\boldsymbol{\gamma})$, $h_l(\boldsymbol{\gamma})$ are Lipschitz continuous. In particular:
$$|h_l(\boldsymbol{\gamma}_1) - h_l(\boldsymbol{\gamma}_2)|, |f(\boldsymbol{\gamma}_1) - f(\boldsymbol{\gamma}_2)| \leq \beta||\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2||$$
whenever $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ satisfy (47), where β is some finite constant that is independent of $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$.

In the special case when $f(\boldsymbol{\gamma})$ and $h_l(\boldsymbol{\gamma})$ are differentiable, then the constant β is determined by the maximum gradient norm over the closed set defined by (47).

C. Analysis of the Generalized Algorithm

We assume optimality of the problem (44)-(46) can be achieved by a (z, Ω)-only policy that has optimal time average penalty vector $\overline{\boldsymbol{x}}$. Define $\boldsymbol{\gamma}^{opt}$ as this optimal time average penalty vector, and define $f^{opt} \triangleq f(\boldsymbol{\gamma}^{opt})$. Note that optimality is still measured with respect to assumed ϕ-forced renewals.

Assumption 3: There is a (z, Ω)-only policy $I^*(t)$ such that for every renewal time t_g we have:

$$\frac{\mathbb{E}\left\{\sum_{\tau=0}^{T^*-1} x_m^*(t_g+\tau)\right\}}{\mathbb{E}\left\{T^*\right\}} = \gamma_m^{opt} \quad \text{for all } m \in \mathcal{M} \tag{57}$$

$$\frac{\mathbb{E}\left\{\sum_{\tau=0}^{T^*-1} d_n^*(t_g+\tau)\right\}}{\mathbb{E}\left\{T^*\right\}} \geq 0 \quad \text{for all } n \in \mathcal{N} \tag{58}$$

where T^* is the renewal interval duration under policy $I^*(t)$, $x_m^*(t)$ and $d_n^*(t)$ correspond to policy $I^*(t)$, and where $\boldsymbol{\gamma}^{opt} = (\gamma_1^{opt}, \ldots, \gamma_M^{opt})$ satisfies $f(\boldsymbol{\gamma}^{opt}) = f^{opt}$ and:

$$h_l(\boldsymbol{\gamma}^{opt}) \leq c_l \ \forall l \in \mathcal{L}$$

We can also impose an additional slackness assumption (similar to Assumption 1) to allow for possible approximate implementations of the weighted stochastic shortest path policy, and to prove strong stability of all queues. However, for simplicity, below we state the performance theorem under the assumption that the exact stochastic shortest path solution to (56) is used every renewal interval. Further, rather than proving strong stability for the virtual queues, we prove a weaker *mean rate stability* result.

Theorem 3: (Performance of the Generalized Weighted Shortest Path Algorithm) Consider any valid definition of a renewal event. Suppose that Assumption 3 holds. For any fixed $V \geq 0$, $\alpha \geq 0$, if the optimal solution to (56) is implemented every renewal interval, then:

(a) For all $n \in \mathcal{N}$, $l \in \mathcal{L}$, $m \in \mathcal{M}$ we have:

$$\lim_{t\rightarrow\infty} \frac{\mathbb{E}\left\{Q_n(t)\right\}}{t} = \lim_{t\rightarrow\infty} \frac{\mathbb{E}\left\{Y_l(t)\right\}}{t} = \lim_{t\rightarrow\infty} \frac{\mathbb{E}\left\{|W_m(t)|\right\}}{t} = 0$$

(b) For all $l \in \mathcal{L}$ we have:

$$\limsup_{t\rightarrow\infty} h_l(\overline{\boldsymbol{x}}(t)) \leq c_l$$

where $\overline{\boldsymbol{x}}(t) \overset{\triangle}{=} \frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\left\{\boldsymbol{x}(\tau)\right\}$.

(c) If renewals are type-2 (so that $\mathbb{E}\left\{T \mid \boldsymbol{\Theta}(t_g)\right\} = 1/\phi$), then the time average cost satisfies:

$$\limsup_{t\rightarrow\infty} f(\overline{\boldsymbol{x}}(t)) \leq f^{opt} + \frac{\phi B_2}{V}$$

(d) If renewals are type-2 and Assumption 4 additionally holds for a slackness value $\epsilon > 0$ (where Assumption 4 is formally stated below), then all queues $Q_n(t)$ for $n \in \mathcal{N}$ are strongly stable, and for any positive integer G:

$$\begin{aligned}\frac{1}{G}\sum_{g=0}^{G-1}\sum_{n\in\mathcal{N}}\mathbb{E}\left\{Q_n(t_g)\right\} \quad &\leq \quad \frac{\phi B_2 + V(f_{max} - f_{min})}{\epsilon} \\ &\quad + \frac{\phi\mathbb{E}\left\{L(\boldsymbol{\Theta}(0))\right\}}{\epsilon G}\end{aligned}$$

The Assumption 4 used in part (d) is given below.

Assumption 4: There is a value $\epsilon > 0$ and a vector $\boldsymbol{\gamma}^* = (\gamma_1^*, \ldots, \gamma_M^*)$ together with a (z, Ω)-only policy $I^*(t)$ (not necessarily the same policy as in Assumption 3) such that for every renewal time t_g:

$$\frac{\mathbb{E}\left\{\sum_{\tau=0}^{T^*-1} x_m^*(t_g+\tau)\right\}}{\mathbb{E}\left\{T^*\right\}} = \gamma_m^* \quad \text{for all } m \in \mathcal{M}$$

$$\frac{\mathbb{E}\left\{\sum_{\tau=0}^{T^*-1} d_n^*(t_g+\tau)\right\}}{\mathbb{E}\left\{T^*\right\}} \geq \epsilon \quad \text{for all } n \in \mathcal{N}$$

where T^* is the renewal interval duration under policy $I^*(t)$, $x_m^*(t)$ and $d_n^*(t)$ correspond to policy $I^*(t)$, and where

$$h_l(\boldsymbol{\gamma}^*) \leq c_l \ \forall l \in \mathcal{L}$$

Proof: Using (54) yields for any renewal time t_g:

$$\begin{aligned}&\Delta_T(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1} f(\boldsymbol{\gamma}(t_g+\tau)) \mid \boldsymbol{\Theta}(t_g)\right\} \leq B_2 \\ &\qquad + D_2(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1} f(\boldsymbol{\gamma}(t_g+\tau)) \mid \boldsymbol{\Theta}(t_g)\right\}\end{aligned}$$

where we assume we are using the stochastic shortest path policy that minimizes (56), and T is the resulting renewal time under this policy. By definition we thus have:

$$\begin{aligned}&\Delta_T(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1} f(\boldsymbol{\gamma}(t_g+\tau)) \mid \boldsymbol{\Theta}(t_g)\right\} \leq B_2 \\ &\qquad + D_2^*(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T^*-1} f(\boldsymbol{\gamma}^*(t_g+\tau)) \mid \boldsymbol{\Theta}(t_g)\right\} \quad (59)\end{aligned}$$

where $D_2^*(\boldsymbol{\Theta}(t_g))$ represents (55) for any other policy $I^*(t)$ and $\boldsymbol{\gamma}^*(t)$, and T^* represents the corresponding renewal duration under this other policy. Now let $I^*(t)$ be the stationary and randomized policy of Assumption 3, and let $\boldsymbol{\gamma}^*(t_g+\tau) = \boldsymbol{\gamma}^{opt}$ for all $t_g + \tau$ in the renewal interval. We thus have (using the definition of $D_2^*(\boldsymbol{\Theta}(t_g))$ in (55)):

$$\begin{aligned}&\Delta_T(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1} f(\boldsymbol{\gamma}(t_g+\tau)) \mid \boldsymbol{\Theta}(t_g)\right\} \\ &\quad \leq \quad B_2 + V\mathbb{E}\left\{T^*\right\} f^{opt} \\ &\qquad - \sum_{n\in\mathcal{N}} Q_n(t_g)\mathbb{E}\left\{\sum_{\tau=0}^{T^*-1} d_n^*(t_g+\tau) \mid \boldsymbol{\Theta}(t_g)\right\} \\ &\qquad - \sum_{m\in\mathcal{M}} W_m(t_g)\mathbb{E}\left\{\sum_{\tau=0}^{T^*-1} [\gamma_m^{opt} - x_m^*(t_g+\tau)] \mid \boldsymbol{\Theta}(t_g)\right\}\end{aligned}$$

where we have used the fact that $h_l(\boldsymbol{\gamma}^{opt}) \leq c_l$ for all $l \in \mathcal{L}$. Plugging in the expressions (57) and (58) yields:

$$\begin{aligned}&\Delta_T(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1} f(\boldsymbol{\gamma}(t_g+\tau)) \mid \boldsymbol{\Theta}(t_g)\right\} \\ &\quad \leq \quad B_2 + V\mathbb{E}\left\{T^*\right\} f^{opt} \qquad (60)\end{aligned}$$

(Proof of part (a)): Because first moments of renewal intervals are bounded by m_1 (by the renewal requirements in Section II-E), and because $f(\boldsymbol{\gamma})$ is bounded, the inequality (60) yields the following for all renewal times t_g:

$$\Delta_T(\boldsymbol{\Theta}(t_g)) \leq B_3$$

where B_3 is a finite positive constant. Taking expectations of the above inequality and summing over $g \in \{0, 1, \ldots, G-1\}$ (where G is any positive integer) yields:

$$\mathbb{E}\left\{L(\boldsymbol{\Theta}(t_G))\right\} \leq GB_3 + \mathbb{E}\left\{L(\boldsymbol{\Theta}(0))\right\}$$

Because the Lyapunov function $L(\boldsymbol{\Theta}(t_G))$ is a sum of squares of queue backlog, we have for any particular queue $Q_n(t_G)$ (for $n \in \mathcal{N}$):

$$\mathbb{E}\left\{Q_n(t_G)\right\}^2 \leq \mathbb{E}\left\{Q_n(t_G)^2\right\} \leq GB_3 + \mathbb{E}\left\{L(\boldsymbol{\Theta}(0))\right\}$$

Taking square roots and dividing by t_G yields:

$$\frac{\mathbb{E}\left\{Q_n(t_G)\right\}}{t_G} \leq \frac{\sqrt{GB_3 + \mathbb{E}\left\{L(\mathbf{\Theta}(0))\right\}}}{t_G}$$

Taking a limit as $G \rightarrow \infty$ and noting that $t_G \geq G$ shows that:

$$\lim_{G\rightarrow\infty} \frac{\mathbb{E}\left\{Q_n(t_G)\right\}}{t_G} = 0$$

While the above limit only samples $Q_n(t)$ at renewal times, it is not difficult to use the fact that mean renewal times are finite to show:

$$\lim_{t\rightarrow\infty} \frac{\mathbb{E}\left\{Q_n(t)\right\}}{t} = 0$$

This derivation holds for all other queues $Y_l(t)$ and $W_m(t)$ in the Lyapunov function, proving part (a).

(Proof of part (b)): Because queues $Y_l(t)$ and $W_m(t)$ are mean rate stable and have dynamics given by (53) and (52), the constraints they enforce hold [18]. In particular, we have:

$$\limsup_{t\rightarrow\infty} \frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\left\{h_l(\boldsymbol{\gamma}(\tau))\right\} \leq c_l \qquad (61)$$

$$\limsup_{t\rightarrow\infty} \frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\left\{x_m(\tau) - \gamma_m(\tau)\right\} = 0 \qquad (62)$$

Now define:

$$\begin{aligned}\overline{x}_m(t) &\triangleq \frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\left\{x_m(\tau)\right\}\\ \overline{\gamma}_m(t) &\triangleq \frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\left\{\gamma_m(\tau)\right\}\end{aligned}$$

Define $\boldsymbol{\alpha}(t)\triangleq\overline{\boldsymbol{\gamma}}(t) - \overline{\boldsymbol{x}}(t)$, and note that (62) implies that $\boldsymbol{\alpha}(t) \rightarrow 0$ as $t \rightarrow \infty$. By Jensen's inequality for the convex function $h_l(\cdot)$, we have:

$$\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\left\{h_l(\boldsymbol{\gamma}(\tau))\right\} \geq h_l(\overline{\boldsymbol{\gamma}}(t)) \geq h_l(\overline{\boldsymbol{x}}(t)) - \beta||\boldsymbol{\alpha}(t)||$$

where we have used the Lipschitz continuity property of $h_l(\cdot)$. Taking a limit as $t \rightarrow \infty$ and using (61) proves part (b).

(Proof of part (c)): Recall that (60) holds for all renewal times t_g for $g \in \{0, 1, 2, \ldots\}$. Taking expectations and summing over $g \in \{0, 1, \ldots, G-1\}$ (for any positive integer G) yields:

$$\mathbb{E}\left\{L(\mathbf{\Theta}(t_G))\right\} - \mathbb{E}\left\{L(\mathbf{\Theta}(0))\right\} + V\mathbb{E}\left\{\sum_{\tau=0}^{t_G-1} f(\boldsymbol{\gamma}(\tau))\right\} \leq GB_2 + VG\mathbb{E}\left\{T^*\right\} f^{opt}$$

Using non-negativity of the Lyapunov function, rearranging terms, and using $\mathbb{E}\left\{T^*\right\} = 1/\phi$ (because we have type-2 renewals) yields:

$$\frac{1}{G/\phi}\mathbb{E}\left\{\sum_{\tau=0}^{t_G-1} f(\boldsymbol{\gamma}(\tau))\right\} \leq f^{opt} + \frac{\phi B_2}{V} + \frac{\phi\mathbb{E}\left\{L(\mathbf{\Theta}(0))\right\}}{VG}$$

Taking a limit as $G \rightarrow \infty$ yields:

$$\limsup_{G\rightarrow\infty} \frac{1}{G/\phi}\mathbb{E}\left\{\sum_{\tau=0}^{t_G-1} f(\boldsymbol{\gamma}(\tau))\right\} \leq f^{opt} + \frac{\phi B_2}{V}$$

However, as in the proof of (82) in Appendix D, the above limit implies:

$$\limsup_{t\rightarrow\infty} \frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\left\{f(\boldsymbol{\gamma}(\tau))\right\} \leq f^{opt} + \frac{\phi B_2}{V} \qquad (63)$$

By using Jensen's inequality on the convex function $f(\boldsymbol{\gamma}(\tau))$ we have for any time t:

$$\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\left\{f(\boldsymbol{\gamma}(\tau))\right\} \geq f(\overline{\boldsymbol{\gamma}}(t)) \geq f(\overline{\boldsymbol{x}}(t)) - \beta||\boldsymbol{\alpha}(t)||$$

Taking a $\limsup$ and combining with (63) completes the proof of part (c).

(Proof of part (d)): Using the policy $I^*(t)$ from Assumption 4, and using $\boldsymbol{\gamma}(t_g + \tau) = \boldsymbol{\gamma}^*$ from Assumption 4 for all τ, from (59) we have:

$$\begin{aligned}&\Delta_T(\mathbf{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1} f(\boldsymbol{\gamma}(t_g + \tau)) \mid \mathbf{\Theta}(t_g)\right\}\\ &\leq \quad B_2 + Vf_{max}/\phi - \sum_{n\in\mathcal{N}} Q_n(t_g)\epsilon/\phi\end{aligned}$$

Because all renewals are geometric with probability ϕ, the second term on the left hand side is at least f_{min}/ϕ and hence:

$$\Delta_T(\mathbf{\Theta}(t_g)) \leq B_2 + V(f_{max} - f_{min})/\phi - \sum_{n\in\mathcal{N}} Q_n(t_g)\epsilon/\phi$$

The above holds for all renewal times t_g for $g \in \{0, 1, 2, \ldots\}$. Taking expectations, summing, and dividing by G yields:

$$\begin{aligned}\frac{\mathbb{E}\left\{L(\mathbf{\Theta}(t_G))\right\} - \mathbb{E}\left\{L(\mathbf{\Theta}(0))\right\}}{G} &\leq B_2 + V(f_{max} - f_{min})/\phi\\ &\quad -\frac{1}{G}\sum_{g=0}^{G-1}\sum_{n\in\mathcal{N}}\mathbb{E}\left\{Q_n(t_g)\right\}\frac{\epsilon}{\phi}\end{aligned}$$

Rearranging terms and using non-negativity of $L(\mathbf{\Theta}(t_G))$ proves part (d). □

VI. Conclusions

We have developed an approach to the constrained Markov Decision problems associated with a small number K of delay-constrained wireless users and a (possibly large) number N of delay-unconstrained queues. Optimization of general penalty functions subject to general penalty constraints and queue stability constraints is treated by reduction to an online (unconstrained) weighted stochastic shortest path problem implemented over variable length frames. This generalizes the class of max-weight network control policies to Markov-modulated networks. The solution to the underlying stochastic shortest path problem has complexity that is geometric in the number of delay-constrained queues K, but polynomial in the number of delay-unconstrained queues N. Explicit bounds on the average backlog in the delay-unconstrained queues were computed and shown to be polynomial in $N + K$. The average size of the virtual queues (which enforce the penalty constraints) were also bounded, which shows that convergence times required to achieve the desired time averages are also polynomial in $N+K$. A Robbins-Monro approximation technique, together with a delayed queue analysis, was shown to provide an efficient online implementation in the case when the system probabilities are not known in advance. The solution technique is general and extends to other Markov modulated networks with general penalties and rewards.

Appendix A — Proof of Lemma 1

We begin with a preliminary lemma (proven at the end of this section).

Lemma 7: For any time t and any integer $T > 0$ we have:

$$\sum_{n\in\mathcal{N}}\left[\left(\sum_{\tau=0}^{T-1}\mu_n(t+\tau)\right)^2 + \left(\sum_{\tau=0}^{T-1}R_n(t+\tau)\right)^2\right] + \sum_{m\in\mathcal{M}}\left(\sum_{\tau=0}^{T-1}|x_m^{av} - x_m(t+\tau)|\right)^2 \leq T^2\sigma^2$$

where σ^2 is defined in (19).

We now use Lemma 7 to prove Lemma 1. From the queue-update equation (1) we have for any queue $n \in \mathcal{N}$ and any renewal time t_g (see, for example, [1] [28]):

$$\begin{aligned} Q_n(t_g+T) &\leq \max\left[Q_n(t_g) - \sum_{\tau=0}^{T-1}\mu_n(t_g+\tau), 0\right] \\ &\quad + \sum_{\tau=0}^{T-1}R_n(t_g+\tau) \end{aligned} \tag{64}$$

Squaring the above equation and using the fact that:

$$\frac{1}{2}(\max[q-\mu, 0] + r)^2 - \frac{q^2}{2} \leq \frac{\mu^2 + r^2}{2} - q(\mu - r) \tag{65}$$

for any non-negative values q, μ, r yields:

$$\begin{aligned} &\tfrac{1}{2}Q_n(t_g+T)^2 - \tfrac{1}{2}Q_n(t_g)^2 \leq \\ &\quad \tfrac{(\sum_{\tau=0}^{T-1}\mu_n(t_g+\tau))^2 + (\sum_{\tau=0}^{T-1}R_n(t_g+\tau))^2}{2} \\ &\quad -Q_n(t_g)\textstyle\sum_{\tau=0}^{T-1}[\mu_n(t_g+\tau) - R_n(t_g+\tau)] \end{aligned} \tag{66}$$

For the $Y_m(t)$ queue, from (15) we have:

$$Y_m(t+1) = \max[Y_m(t) - \alpha_m(t), 0]$$

where we define $\alpha_m(t) \triangleq x_m^{av} - x_m(t)$. Similar to (66), we have the following lemma (proven at the end of this section).

Lemma 8:

$$\begin{aligned} \frac{1}{2}Y_m(t_g+T)^2 &\leq \frac{1}{2}Y_m(t_g)^2 - Y_m(t_g)\sum_{\tau=0}^{T-1}\alpha_m(t_g+\tau) \\ &\quad + \frac{1}{2}\left(\sum_{\tau=0}^{T-1}|\alpha_m(t_g+\tau)|\right)^2 \end{aligned} \tag{67}$$

Combining (66) and (67), using Lemma 7, and taking conditional expectations yields:

$$\begin{aligned} \Delta_T(\boldsymbol{\Theta}(t_g)) \leq B - \sum_{m\in\mathcal{M}} Y_m(t_g)\mathbb{E}\left\{\sum_{\tau=0}^{T-1}\alpha_m(t_g+\tau) \mid \boldsymbol{\Theta}(t_g)\right\} \\ - \sum_{n\in\mathcal{N}} Q_n(t_g)\mathbb{E}\left\{\sum_{\tau=0}^{T-1}[\mu_n(t_g+\tau) - R_n(t_g+\tau)] \mid \boldsymbol{\Theta}(t_g)\right\} \end{aligned}$$

This proves Lemma 1.

It remains only to prove Lemmas 7 and 8.

Proof: (Lemma 7) Define the following vector $\boldsymbol{\beta}(\tau)$:

$$\boldsymbol{\beta}(\tau) \triangleq [(\mu_n(\tau))_{n\in\mathcal{N}}, (R_n(\tau))_{n\in\mathcal{N}}, (|\alpha_m(\tau)|)_{m\in\mathcal{M}}]$$

Denote by $\mathcal{A}$ the set of all possible values that can be achieved by the vector $\boldsymbol{\beta}(\tau)$ for a single timeslot τ (considering all possible control actions and all possible random outcomes $\Omega(\tau)$). Note that for any time t and any integer $T > 0$:

$$\frac{1}{T}\sum_{\tau=0}^{T-1}\boldsymbol{\beta}(t+\tau) \in Conv(\mathcal{A}) \tag{68}$$

where $Conv(\mathcal{A})$ is the convex hull of the set $\mathcal{A}$. Now define the following convex function:

$$f(\boldsymbol{\beta}(\tau)) \triangleq \sum_{n\in\mathcal{N}}[\mu_n(\tau)^2 + R_n(\tau)^2] + \sum_{m\in\mathcal{M}}|\alpha_m(\tau)|^2$$

Note also that $f(T\boldsymbol{\beta}(t)) = T^2 f(\boldsymbol{\beta}(t))$. Thus:

$$\begin{aligned} f\left(\sum_{\tau=0}^{T-1}\boldsymbol{\beta}(t+\tau)\right) &= T^2 f\left(\frac{1}{T}\sum_{\tau=0}^{T-1}\boldsymbol{\beta}(t+\tau)\right) \\ &\leq T^2 \sup_{\boldsymbol{\beta}\in Conv(\mathcal{A})} f(\boldsymbol{\beta}) \qquad (69) \\ &= T^2 \sup_{\boldsymbol{\beta}\in\mathcal{A}} f(\boldsymbol{\beta}) \qquad (70) \\ &\leq T^2\sigma^2 \qquad (71) \end{aligned}$$

where (69) follows by (68), (70) follows because the supremum of a convex function over the convex hull of a set $\mathcal{A}$ is the same as the supremum over the set $\mathcal{A}$ itself, and (71) follows by definition of σ^2 in (19). □

Proof: (Lemma 8) Recall that $Y_m(t+1) = \max[Y_m(t) - \alpha_m(t), 0]$. Define:

$$\begin{aligned} \alpha_m^{pos}(t) &\triangleq \max[\alpha_m(t), 0] \\ \alpha_m^{neg}(t) &\triangleq -\min[\alpha_m(t), 0] \end{aligned}$$

so that $\alpha_m^{pos}(t) \geq 0$, $\alpha_m^{neg}(t) \geq 0$, and:

$$\begin{aligned} \alpha_m(t) &= \alpha_m^{pos}(t) - \alpha_m^{neg}(t) \\ |\alpha_m(t)| &= \alpha_m^{pos}(t) + \alpha_m^{neg}(t) \end{aligned}$$

Similar to (64), it is not difficult to show that:

$$\begin{aligned} Y_m(t+T) \leq \max\left[Y_m(t) - \sum_{\tau=0}^{T-1}\alpha_m^{pos}(t+\tau), 0\right] \\ + \sum_{\tau=0}^{T-1}\alpha_m^{neg}(t+\tau) \end{aligned}$$

Therefore, using (65):

$$\begin{aligned} \frac{1}{2}Y_m(t+T)^2 &\leq \frac{1}{2}Y_m(t)^2 - Y_m(t)\sum_{\tau=0}^{T-1}\alpha_m(t+\tau) \\ &\quad + \frac{1}{2}\left(\sum_{\tau=0}^{T-1}\alpha_m^{pos}(t+\tau)\right)^2 \\ &\quad + \frac{1}{2}\left(\sum_{\tau=0}^{T-1}\alpha_m^{neg}(t+\tau)\right)^2 \end{aligned}$$

The result of inequality (67) follows from the above together with the fact that:

$$\begin{aligned}
&\frac{1}{2}\left(\sum_{\tau=0}^{T-1}\alpha_m^{pos}(t+\tau)\right)^2 + \frac{1}{2}\left(\sum_{\tau=0}^{T-1}\alpha_m^{neg}(t+\tau)\right)^2 \\
&\leq \quad \frac{1}{2}\left(\sum_{\tau=0}^{T-1}[\alpha_m^{pos}(t+\tau)+\alpha_m^{neg}(t+\tau)]\right)^2 \\
&= \quad \frac{1}{2}\left(\sum_{\tau=0}^{T-1}|\alpha_m(t+\tau)|\right)^2
\end{aligned}$$

□

Appendix B — Proof that (23) Implies Strong Stability

The inequality (23) implies that for every integer $G \geq 1$:

$$\frac{1}{G}\sum_{g=0}^{G-1}\mathbb{E}\left\{Q_n(t_g)\right\} \leq q \tag{72}$$

for some finite constant q. Here we show that this implies $\overline{Q}_n < \infty$ (the same reasoning also shows that $\overline{Y}_m < \infty$). Note from (1) that the queue $Q_n(t)$ can increase by at most a finite constant r on any slot (where $r = R_n^{max}$). Thus, for any renewal time t_g (with renewal size T_g) and for any $\tau \in \{0, 1, \ldots, T_g - 1\}$ we have:

$$Q_n(t_g+\tau) \leq Q_n(t_g) + r\tau$$

Using the above inequality yields:

$$\begin{aligned}
\mathbb{E}\left\{\sum_{\tau=0}^{t_G-1}Q_n(\tau)\right\} \quad \leq \quad & \sum_{g=0}^{G-1}\mathbb{E}\left\{T_gQ_n(t_g)\right\} + \\
& \sum_{g=0}^{G-1}\frac{r\mathbb{E}\left\{T_g(T_g-1)\right\}}{2}
\end{aligned} \tag{73}$$

Now recall that m_1 and m_2 are finite bounds on the first and second moment of any renewal time T_g, and these bounds hold regardless of past events before this renewal interval and hence regardless of the value of $Q_n(t_g)$. We thus have:

$$\begin{aligned}
\mathbb{E}\left\{T_gQ_n(t_g)\right\} &= \mathbb{E}\left\{\mathbb{E}\left\{T_gQ_n(t_g) \mid Q_n(t_g)\right\}\right\} \\
&= \mathbb{E}\left\{Q_n(t_g)\mathbb{E}\left\{T_g \mid Q_n(t_g)\right\}\right\} \\
&\leq \mathbb{E}\left\{Q_n(t_g)m_1\right\} = m_1\mathbb{E}\left\{Q_n(t_g)\right\}
\end{aligned}$$

Using this in (73) yields:

$$\mathbb{E}\left\{\sum_{\tau=0}^{t_G-1}Q_n(\tau)\right\} \quad \leq \quad m_1\sum_{g=0}^{G-1}\mathbb{E}\left\{Q_n(t_g)\right\} + \frac{Grm_2}{2}$$

Dividing by G and using (72) yields:

$$\frac{1}{G}\mathbb{E}\left\{\sum_{\tau=0}^{t_G-1}Q_n(\tau)\right\} \leq m_1q + \frac{rm_2}{2}$$

Now note that all renewal intervals are at least one slot, so that $G \leq t_G$. Using this with the fact that queue values are non-negative yields:

$$\frac{1}{G}\sum_{\tau=0}^{G-1}\mathbb{E}\left\{Q_n(\tau)\right\} \leq \frac{1}{G}\mathbb{E}\left\{\sum_{\tau=0}^{t_G-1}Q_n(\tau)\right\} \leq m_1q + \frac{rm_2}{2}$$

The above holds for all positive integers G. Taking a limit as $G \rightarrow \infty$ shows that $\overline{Q}_n < \infty$ and proves the result. Strong stability of the queues $Q_n(t)$ and $Y_m(t)$, together with the fact that these queues have finite maximum transmission rates, implies the time average constraints (3)-(4) are satisfied [18].

Appendix C — Proof of Theorem 2

Here we prove Theorem 2. Let t be a renewal time (for type-2 renewals), and let T be the renewal duration under the implemented policy. From (28) and (17) we have:

$$\begin{aligned}
&\Delta_T(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1}x_0(t_g+\tau) \mid \boldsymbol{\Theta}(t_g)\right\} \leq B + C \\
&\qquad + D^{ssp}(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1}x_0^{ssp}(t_g+\tau) \mid \boldsymbol{\Theta}(t_g)\right\} \\
&\qquad + \delta\sum_{n\in\mathcal{N}}Q_n(t_g) + \delta\sum_{m\in\mathcal{M}}Y_m(t_g) + V\delta
\end{aligned} \tag{74}$$

By definition of the weighted stochastic shortest path policy, we have:

$$\begin{aligned}
&D^{ssp}(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1}x_0^{ssp}(t_g+\tau) \mid \boldsymbol{\Theta}(t_g)\right\} \leq \\
&\qquad D^{*}(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1}x_0^{*}(t_g+\tau) \mid \boldsymbol{\Theta}(t_g)\right\}
\end{aligned} \tag{75}$$

where $D^*(\boldsymbol{\Theta}(t_g))$ and $x_0^*(t_g+\tau)$ correspond to any other control policy $I^*(t)$ that could be implemented over the renewal interval (note that the corresponding renewal interval size T does not change, so that $T^* = T$, because type-2 renewal events do not depend on control decisions). Using (75) in (74) together with the definition of $D^*(\boldsymbol{\Theta}(t_g))$ given in (18) yields:

$$\begin{aligned}
&\Delta_T(\boldsymbol{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1}x_0(t_g+\tau) \mid \boldsymbol{\Theta}(t_g)\right\} \leq B + C \\
&\qquad + V\delta + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1}x_0^{*}(t_g+\tau) \mid \boldsymbol{\Theta}(t_g)\right\} \\
&\qquad - \sum_{n\in\mathcal{N}}Q_n(t_g)\mathbb{E}\left\{\sum_{\tau=0}^{T-1}d_n^{*}(t_g+\tau) - \delta \mid \boldsymbol{\Theta}(t_g)\right\} \\
&- \sum_{m\in\mathcal{M}}Y_m(t_g)\mathbb{E}\left\{-\delta + \sum_{\tau=0}^{T-1}[x_m^{av} - x_m^{*}(t_g+\tau)] \mid \boldsymbol{\Theta}(t_g)\right\}
\end{aligned} \tag{76}$$

where we have used the following notation:

$$\begin{aligned}
x_m^*(\tau) &= \hat{x}_m(I^*(\tau), \Omega(\tau), z^*(\tau)) \\
d_n^*(\tau) &= \hat{d}_n(I^*(\tau), \Omega(\tau), z^*(\tau))
\end{aligned}$$

Now choose $I^*(t)$ as the policy of Assumption 1 that yields (10)-(11). Plugging (10)-(11) directly into (76) and using the fact that $T = T^*$ and $\mathbb{E}\{T \mid \mathbf{\Theta}\} = 1/\phi$ yields:

$$\Delta_T(\mathbf{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1} x_0(t_g+\tau) \mid \mathbf{\Theta}(t_g)\right\} \leq B + C + V\delta + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1} x_0^*(t_g+\tau) \mid \mathbf{\Theta}(t_g)\right\} - \sum_{n\in\mathcal{N}} Q_n(t_g)(\epsilon/\phi - \delta) - \sum_{m\in\mathcal{M}} Y_m(t_g)(\epsilon/\phi - \delta) \quad (77)$$

Using the bounds x_0^{min} and x_0^{max} in the above inequality and rearranging terms yields:

$$\Delta_T(\mathbf{\Theta}(t_g)) + \sum_{n\in\mathcal{N}} Q_n(t_g)(\epsilon/\phi-\delta) + \sum_{m\in\mathcal{M}} Y_m(t_g)(\epsilon/\phi-\delta) \leq B + C + V(\delta + (x_0^{max} - x_0^{min})/\phi)$$

Taking expectations, summing over all renewal events t_g (for $g \in \{0, 1, 2, \ldots, G-1\}$ and $t_0 = 0$) and using telescoping sums, non-negativity of $L(\cdot)$, together with the fact that $L(\mathbf{\Theta}(0)) = 0$ (as in the proof of Theorem 1) yields:

$$\frac{1}{G}\sum_{g=0}^{G-1}\left[\sum_{n\in\mathcal{N}} \mathbb{E}\{Q_n(t_g)\} + \sum_{m\in\mathcal{M}} \mathbb{E}\{Y_m(t_g)\}\right] \leq \frac{B + C + V(\delta + (x_0^{max} - x_0^{min})/\phi)}{\epsilon/\phi - \delta}$$

This proves (29).

To prove (30) consider again the drift inequality (76), but now plug in the following policy $I^*(t)$: Define probability $\theta \triangleq \delta\phi/\epsilon$. This is a valid probability because $\epsilon > \phi\delta$ by assumption. At each time t_g that marks the beginning of a renewal, independently flip a biased coin with probabilities θ and $1-\theta$, and carry out one of the two following policies for the full duration of the renewal interval:

- With probability θ: Use the stationary randomized policy from Assumption 1 (which we shall call $I_1^*(t)$), for the duration of the renewal interval, which yields (10)-(11).
- With probability $1-\theta$: Use the stationary randomized policy from Assumption 2 (here denoted $I_2^*(t)$), for the duration of the renewal interval, which yields (12)-(14).

With this policy $I^*(t)$, from (12) we have (recall $T^* = T$):

$$\mathbb{E}\left\{\sum_{\tau=0}^{T-1} x_0^*(t_g+\tau)\right\} \leq \frac{\theta x_0^{max} + (1-\theta)x_0^{opt}}{\phi} \quad (78)$$

We also have from (10)-(11) and (13)-(14):

$$\mathbb{E}\left\{\sum_{\tau=0}^{T-1} x_m^*(t_g+\tau)\right\} \leq \frac{\theta(x_m^{av} - \epsilon) + (1-\theta)x_m^{av}}{\phi} \quad (79)$$

$$\mathbb{E}\left\{\sum_{\tau=0}^{T-1} d_n^*(t_g+\tau)\right\} \geq \frac{\theta\epsilon}{\phi} \quad (80)$$

Plugging (78)-(80) into (76) and using the definition of $\theta = \delta\phi/\epsilon$ yields:

$$\Delta_T(\mathbf{\Theta}(t_g)) + V\mathbb{E}\left\{\sum_{\tau=0}^{T-1} x_0(t_g+\tau) \mid \mathbf{\Theta}(t_g)\right\} \leq B + C + V\delta + V\frac{1}{\phi}[\theta x_0^{max} + (1-\theta)x_0^{opt}]$$

The above holds for all times t_g that mark the beginning of renewal intervals. Defining $T = T_g$ (the duration of the gth renewal interval) and taking expectations of the above inequality yields:

$$\mathbb{E}\left\{L(\mathbf{\Theta}(t_g+T_g)) - L(\mathbf{\Theta}(t_g)) + V\sum_{\tau=0}^{T_g-1} x_0(t_g+\tau)\right\} \leq B + C + V\delta + \frac{V\delta(x_0^{max} - x_0^{opt})}{\epsilon} + \frac{Vx_0^{opt}}{\phi}$$

Summing over $g \in \{0, \ldots, G-1\}$, dividing by VG/ϕ, and using the fact that $L(\mathbf{\Theta}(t_G)) \geq 0$ and $L(\mathbf{\Theta}(0)) = 0$ yields for any positive integer G:

$$\frac{\mathbb{E}\left\{\sum_{\tau=0}^{t_G-1} x_0(\tau)\right\}}{G/\phi} \leq x_0^{opt} \frac{(B+C)\phi}{V} + \delta\phi + \frac{\delta\phi}{\epsilon}(x_0^{max} - x_0^{opt}) \quad (81)$$

Because renewal intervals $\{T_g\}_{g=1}^{\infty}$ are i.i.d. geometric random variables with $\mathbb{E}\{T_g\} = 1/\phi$, we have by the Law of Large Numbers:

$$\frac{t_G}{G} = \frac{\sum_{g=0}^{G-1} T_g}{G} \rightarrow 1/\phi \text{ with prob. 1}$$

Using this together with the fact that $x_0(\tau)$ is upper and lower bounded for all τ, it can be shown that (see Appendix D):

$$\limsup_{G\rightarrow\infty} \frac{\mathbb{E}\left\{\sum_{\tau=0}^{t_G-1} x_0(\tau)\right\}}{G/\phi} = \limsup_{t\rightarrow\infty} \frac{1}{t}\sum_{\tau=0}^{t-1} \mathbb{E}\{x_0(\tau)\} \quad (82)$$

Using this fact in (81) proves (30).

Appendix D — Proof of (82)

First consider the case when $x_0^{min} \geq 0$, so that $0 \leq x_0(t) \leq x_0^{max}$ for all t. Let $G(t)$ be the number of renewal events that have occurred up to time t (not counting the renewal at time 0). Then $G(t)/t \rightarrow \phi$ with probability 1. Fix a value $\epsilon > 0$ such that $0 < \epsilon < \phi$. Define the following event $\chi(t)$:

$$\chi(t) \triangleq \left\{\frac{G(t)}{t} < \phi + \epsilon\right\}$$

Define $\chi^c(t)$ as the opposite event. Then $Pr[\chi^c(t)] \rightarrow 0$ as $t \rightarrow \infty$. If $\chi(t)$ is true, then $G(t) < \lceil(\phi+\epsilon)t\rceil$ and so:

$$t < t_{\lceil(\phi+\epsilon)t\rceil} \text{ whenever } \chi(t) \text{ is true}$$

where we recall that t_G is the time of the Gth renewal event. Now for any time t we have:

$$\begin{aligned}
\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\left\{x_0(\tau)\right\} &= \mathbb{E}\left\{\frac{1}{t}\sum_{\tau=0}^{t-1}x_0(\tau)\mid\chi(t)\right\}Pr[\chi(t)] \\
&\quad +\mathbb{E}\left\{\frac{1}{t}\sum_{\tau=0}^{t-1}x_0(\tau)\mid\chi^c(t)\right\}Pr[\chi^c(t)] \\
&\leq \mathbb{E}\left\{\frac{1}{t}\sum_{\tau=0}^{t_{\lceil(\phi+\epsilon)t\rceil}-1}x_0(\tau)\mid\chi(t)\right\}Pr[\chi(t)] \\
&\quad +x_0^{max}Pr[\chi^c(t)] \\
&\leq \frac{1}{t}\mathbb{E}\left\{\sum_{\tau=0}^{t_{\lceil(\phi+\epsilon)t\rceil}-1}x_0(\tau)\right\} \\
&\quad +x_0^{max}Pr[\chi^c(t)]
\end{aligned}$$

where the final inequality holds because we have added the non-negative term:

$$\frac{1}{t}\mathbb{E}\left\{\sum_{\tau=0}^{t_{\lceil(\phi+\epsilon)t\rceil}-1}x_0(\tau)\mid\chi^c(t)\right\}Pr[\chi^c(t)]$$

Therefore:

$$\begin{aligned}
&\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\left\{x_0(\tau)\right\} \\
&\leq \frac{\lceil(\phi+\epsilon)t\rceil}{t}\frac{1}{\lceil(\phi+\epsilon)t\rceil}\mathbb{E}\left\{\sum_{\tau=0}^{t_{\lceil(\phi+\epsilon)t\rceil}-1}x_0(\tau)\right\} \\
&\quad +x_0^{max}Pr[\chi^c(t)]
\end{aligned}$$

Taking limits yields:

$$\limsup_{t\rightarrow\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\left\{x_0(\tau)\right\}\leq(\phi+\epsilon)\limsup_{G\rightarrow\infty}\frac{1}{G}\mathbb{E}\left\{\sum_{\tau=0}^{t_G-1}x_0(\tau)\right\}$$

The above holds for all ϵ such that $0<\epsilon<\phi$. Taking a limit as $\epsilon\rightarrow 0$ yields:

$$\limsup_{t\rightarrow\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\left\{x_0(\tau)\right\}\leq\phi\limsup_{G\rightarrow\infty}\frac{1}{G}\mathbb{E}\left\{\sum_{\tau=0}^{t_G-1}x_0(\tau)\right\}$$

The reverse inequality can be proven similarly. This establishes (82) for the case when $x_0^{min}\geq 0$.

For the case $x_0^{min}<0$, we can define $\tilde{x}_0(\tau)\triangleq x_0(\tau)-x_{min}$. Then we have $0\leq\tilde{x}_0(\tau)\leq x_0^{max}-x_0^{min}$ for all τ. It follows that:

$$\limsup_{t\rightarrow\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\left\{\tilde{x}_0(\tau)\right\}=\phi\limsup_{G\rightarrow\infty}\frac{1}{G}\mathbb{E}\left\{\sum_{\tau=0}^{t_G-1}\tilde{x}_0(\tau)\right\}$$

Adding x_0^{min} to both sides of the above equality yields the result of (82).

Appendix E — Convergence Proofs

Here we prove Lemmas 2, 3, and 4 of Section IV-A. We have two preliminary lemmas.

Lemma 9: If P is a transition probability matrix (with rows given by probabilities that sum to 1), then for any random column vector $\boldsymbol{X}$ with dimension equal to the number of columns of P, we have:

$$||P\boldsymbol{X}||_e\leq||\boldsymbol{X}||_e$$

Proof: For every row i of $P\boldsymbol{X}$ we have by Jensen's inequality:

$$(\textstyle\sum_j P_{ij}X_j)^2\leq\sum_j P_{ij}X_j^2$$

and hence:

$$\mathbb{E}\left\{(\textstyle\sum_j P_{ij}X_j)^2\right\}\leq\sum_j P_{ij}\mathbb{E}\left\{X_j^2\right\}\leq\max_j\mathbb{E}\left\{X_j^2\right\}$$

Thus:

$$\max_i\mathbb{E}\left\{(\textstyle\sum_j P_{ij}X_j)^2\right\}\leq\max_j\mathbb{E}\left\{X_j^2\right\}$$

The left hand side of the above inequality is $||P\boldsymbol{X}||_e^2$, and the right hand side is $||\boldsymbol{X}||_e^2$, proving the lemma. □

We now show that the map Ψ from (35) is a contraction with respect to the norm $||\cdot||_e$.

Lemma 10: For any two random vectors $\boldsymbol{J}_1$ and $\boldsymbol{J}_2$, we have:

$$||\Psi\boldsymbol{J}_1-\Psi\boldsymbol{J}_2||_e\leq(1-\phi)||\boldsymbol{J}_1-\boldsymbol{J}_2||_e$$

Therefore, letting $\boldsymbol{J}^*$ denote the unique fixed point solution to (34) (satisfying $\boldsymbol{J}^*=\Psi\boldsymbol{J}^*$), then for any random vector $\boldsymbol{J}$ we have:

$$||\Psi\boldsymbol{J}-\boldsymbol{J}^*||_e\leq(1-\phi)||\boldsymbol{J}-\boldsymbol{J}^*||_e$$

Proof: From (35) we have:

$$\begin{aligned}
\Psi\boldsymbol{J}_1 = \phi\mathbb{E}&\left\{\min_{I\in\mathcal{I}_{[\omega(t),1],\boldsymbol{z}}}\boldsymbol{c}_{\boldsymbol{\Theta}}^{(1)}(I,\omega(t))\right\}+ \\
&(1-\phi)\mathbb{E}\left\{c_{\boldsymbol{\Theta}}^{(0)}(I_1,\omega(t))+P^{(0)}(I_1,\omega(t))\boldsymbol{J}_1\mid\boldsymbol{J}_1,\boldsymbol{J}_2\right\}
\end{aligned}$$

where I_1 represents the policy that minimizes the expectation in the final term (corresponding to vector $\boldsymbol{J}_1$), and the expectation is with respect to the random $\omega(t)$. The additional conditioning on $\boldsymbol{J}_2$ does not change anything and is done to facilitate the next few steps. Similarly:

$$\begin{aligned}
\Psi\boldsymbol{J}_2 = \phi\mathbb{E}&\left\{\min_{I\in\mathcal{I}_{[\omega(t),1],\boldsymbol{z}}}\boldsymbol{c}_{\boldsymbol{\Theta}}^{(1)}(I,\omega(t))\right\}+ \\
&(1-\phi)\mathbb{E}\left\{c_{\boldsymbol{\Theta}}^{(0)}(I_2,\omega(t))+P^{(0)}(I_2,\omega(t))\boldsymbol{J}_2\mid\boldsymbol{J}_1,\boldsymbol{J}_2\right\}
\end{aligned}$$

where I_2 is the policy that minimizes the expectation in the final term (corresponding to vector $\boldsymbol{J}_2$). We thus have:

$$\begin{aligned}
\Psi\boldsymbol{J}_2 &\leq \phi\mathbb{E}\left\{\min_{I\in\mathcal{I}_{[\omega(t),1],\boldsymbol{z}}}\boldsymbol{c}_{\boldsymbol{\Theta}}^{(1)}(I,\omega(t))\right\}+ \\
&\quad (1-\phi)\mathbb{E}\left\{c_{\boldsymbol{\Theta}}^{(0)}(I_1,\omega(t))+P^{(0)}(I_1,\omega(t))\boldsymbol{J}_2\mid\boldsymbol{J}_1,\boldsymbol{J}_2\right\} \\
&= \phi\mathbb{E}\left\{\min_{I\in\mathcal{I}_{[\omega(t),1],\boldsymbol{z}}}\boldsymbol{c}_{\boldsymbol{\Theta}}^{(1)}(I,\omega(t))\right\}+ \\
&\quad (1-\phi)\mathbb{E}\left\{c_{\boldsymbol{\Theta}}^{(0)}(I_1,\omega(t))+P^{(0)}(I_1,\omega(t))\boldsymbol{J}_1\mid\boldsymbol{J}_1,\boldsymbol{J}_2\right\} \\
&\quad +(1-\phi)\mathbb{E}\left\{P^{(0)}(I_1,\omega(t))\right\}(\boldsymbol{J}_2-\boldsymbol{J}_1) \\
&= \Psi\boldsymbol{J}_1+(1-\phi)\mathbb{E}\left\{P^{(0)}(I_1,\omega(t))\right\}(\boldsymbol{J}_2-\boldsymbol{J}_1)
\end{aligned}$$

Therefore we conclude:

$$\Psi\boldsymbol{J}_2-\Psi\boldsymbol{J}_1\leq(1-\phi)P_1(\boldsymbol{J}_2-\boldsymbol{J}_1) \tag{83}$$

where we define $P_1 \triangleq \mathbb{E}\left\{P^{(0)}(I_1, \omega(t))\right\}$, and note that each row of P_1 consists of probabilities that sum to 1 (as it is the expectation of matrices with that property). Similarly, by swapping the roles of $\boldsymbol{J}_1$ and $\boldsymbol{J}_2$, we have:

$$\Psi \boldsymbol{J}_1 - \Psi \boldsymbol{J}_2 \leq (1-\phi) P_2 (\boldsymbol{J}_1 - \boldsymbol{J}_2)$$

where P_2 is a transition probability matrix (with rows that sum to 1) defined $P_2 \triangleq \mathbb{E}\left\{P^{(0)}(I_2, \omega(t))\right\}$. It follows that:

$$\begin{aligned} &||\Psi \boldsymbol{J}_2 - \Psi \boldsymbol{J}_1||_e \\ &\leq \quad (1-\phi)\max[||P_1(\boldsymbol{J}_2 - \boldsymbol{J}_1)||_e, ||P_2(\boldsymbol{J}_1 - \boldsymbol{J}_2)||_e] \\ &\leq \quad (1-\phi)||\boldsymbol{J}_2 - \boldsymbol{J}_1||_e \end{aligned}$$

where the final inequality follows by Lemma 9. □

Now consider the iteration:

$$\boldsymbol{J}_{b+1} = \gamma \tilde{\Psi} \boldsymbol{J}_b + (1-\gamma)\boldsymbol{J}_b \tag{84}$$

where γ satisfies $0 < \gamma < 1$, and where:

$$\tilde{\Psi}\boldsymbol{J}_b = \Psi \boldsymbol{J}_b + \boldsymbol{\eta}_b$$

where Ψ is the map of (35), and $\{\boldsymbol{\eta}_b\}_{b=1}^{\infty}$ is a sequence of zero mean vector random variables, where each entry of $\boldsymbol{\eta}_b$ is uncorrelated with any deterministic function of $\boldsymbol{J}_b$. We show that $||\boldsymbol{J}_b||_d$ and $||\boldsymbol{\eta}_b||_d$ are deterministically bounded.

Lemma 2: Define $J_{max} \triangleq c_{max}/\phi$. If $||\boldsymbol{J}_0||_d \leq J_{max}$, then:

(a) For all $b \in \{0, 1, 2, \ldots\}$ we have:

$$||\boldsymbol{J}_b||_d \leq J_{max}$$

(b) There are finite constants η_{min} and η_{max} such that $\eta_{min} \leq \eta_b[i] \leq \eta_{max}$ for all iterations b and all entries i. Further, for all b, if batches of size $L \geq 1$ are used, then:

$$||\boldsymbol{\eta}_b||_e^2 \quad \leq \quad \frac{|\eta_{min}\eta_{max}|}{L} \leq \frac{4(c_{max} + (1-\phi)J_{max})^2}{L}$$

Proof: Suppose that $||\boldsymbol{J}_b||_d \leq J_{max}$ for some iteration $b \geq 0$ (it holds by assumption for $b = 0$). We show that it also holds for $b+1$. By the update equations (35) and (36), it is not difficult to show that:

$$\max[||\tilde{\Psi}\boldsymbol{J}_b||_d, ||\Psi \boldsymbol{J}_b||_d] \leq c_{max} + (1-\phi)J_{max}$$

Thus:

$$\begin{aligned} ||\boldsymbol{J}_{b+1}||_d \quad &\leq \quad \gamma[c_{max} + (1-\phi)J_{max}] + (1-\gamma)J_{max} \\ &= \quad \gamma c_{max} + (1-\phi\gamma)J_{max} \\ &= \quad \gamma\phi J_{max} + (1-\phi\gamma)J_{max} = J_{max} \end{aligned}$$

This proves part (a). Part (b) follows because:

$$||\boldsymbol{\eta}_b||_d = ||\tilde{\Psi}\boldsymbol{J}_b - \Psi \boldsymbol{J}_b||_d \leq 2c_{max} + 2(1-\phi)J_{max}$$

Further, if $\boldsymbol{\eta}_b$ is based only on one sample (so that $L = 1$), then its variance is bounded by $|\eta_{min}\eta_{max}|$ (see Lemma 11 in Appendix F). If it is based on L i.i.d. samples, then the variance is reduced by a factor of L. □

Lemma 3: Let $\boldsymbol{J}_b$ be the bth iteration of (39), starting with some initial vector $\boldsymbol{J}_0$ with $||\boldsymbol{J}_0||_e \leq J_{max}$. Assume that for all b we have $||\boldsymbol{\eta}_b||_e^2 \leq \sigma^2$ for some finite constant σ^2 (as in Lemma 2 with $\sigma^2 = |\eta_{min}\eta_{max}|/L$). Let $\boldsymbol{J}^*$ be the optimal solution to (34), satisfying $\Psi \boldsymbol{J}^* = \boldsymbol{J}^*$. Then:

(a) Every one-step iteration satisfies (for integers $b \geq 0$):

$$||\boldsymbol{J}_{b+1} - \boldsymbol{J}^*||_e^2 \leq (1-\phi\gamma)^2 ||\boldsymbol{J}_b - \boldsymbol{J}^*||_e^2 + \gamma^2 ||\boldsymbol{\eta}_b||_e^2$$

(b) After b iterations we have:

$$\begin{aligned} ||\boldsymbol{J}_b - \boldsymbol{J}^*||_e^2 \quad \leq \quad & (1-\phi\gamma)^{2b}||\boldsymbol{J}_0 - \boldsymbol{J}^*||_e^2 \\ & + \frac{\gamma\sigma^2(1-(1-\phi\gamma)^{2b})}{\phi(2-\phi\gamma)} \end{aligned}$$

(c) In the limit, we have:

$$\lim_{b\rightarrow\infty} ||\boldsymbol{J}_b - \boldsymbol{J}^*||_e^2 \leq \frac{\gamma\sigma^2}{\phi(2-\phi\gamma)}$$

Proof: To prove part (a), we have:

$$\begin{aligned} &||\boldsymbol{J}_{b+1} - \boldsymbol{J}^*||_e^2 \\ &= \quad ||\gamma\Psi\boldsymbol{J}_b + \gamma\boldsymbol{\eta}_b + (1-\gamma)\boldsymbol{J}_b - \boldsymbol{J}^*||_e^2 \\ &= \quad \max_i \mathbb{E}\left\{(\gamma\Psi\boldsymbol{J}_b + \gamma\boldsymbol{\eta}_b + (1-\gamma)\boldsymbol{J}_b - \boldsymbol{J}^*)[i]^2\right\} \end{aligned}$$

where the final term represents the ith entry of the random vector:

$$\gamma\Psi\boldsymbol{J}_b + \gamma\boldsymbol{\eta}_b + (1-\gamma)\boldsymbol{J}_b - \boldsymbol{J}^*$$

However, each entry i of the $\boldsymbol{\eta}_b$ vector is zero mean and uncorrelated with any deterministic function of $\boldsymbol{J}_b$. Thus, the second moment of the ith entry is equal to the sum of the second moment of the $\gamma\boldsymbol{\eta}_b[i]$ component and the second moment of the remaining components. Thus:

$$\begin{aligned} &||\boldsymbol{J}_{b+1} - \boldsymbol{J}^*||_e^2 \\ &\leq \quad ||\gamma\Psi\boldsymbol{J}_b + (1-\gamma)\boldsymbol{J}_b - \boldsymbol{J}^*||_e^2 + \gamma^2||\boldsymbol{\eta}_b||_e^2 \end{aligned} \tag{85}$$

However:

$$\begin{aligned} &||\gamma\Psi\boldsymbol{J}_b + (1-\gamma)\boldsymbol{J}_b - \boldsymbol{J}^*||_e^2 \\ &= \quad ||\gamma\Psi(\boldsymbol{J}_b - \boldsymbol{J}^*) + (1-\gamma)(\boldsymbol{J}_b - \boldsymbol{J}^*)||_e^2 && (86) \\ &\leq \quad (\gamma||\Psi(\boldsymbol{J}_b - \boldsymbol{J}^*)||_e + (1-\gamma)||\boldsymbol{J}_b - \boldsymbol{J}^*||_e)^2 && (87) \\ &\leq \quad (\gamma(1-\phi)||\boldsymbol{J}_b - \boldsymbol{J}^*||_e + (1-\gamma)||\boldsymbol{J}_b - \boldsymbol{J}^*||_e)^2 && (88) \\ &= \quad (1-\phi\gamma)^2||\boldsymbol{J}_b - \boldsymbol{J}^*||_e^2 \end{aligned}$$

where (86) follows because $\Psi\boldsymbol{J}^* = \boldsymbol{J}^*$, (87) uses the triangle inequality, and (88) uses the contraction property of Ψ from Lemma 10. Combining the above with (85) establishes part (a).

To prove part (b), we have from repetitions of the iteration in part (a):

$$\begin{aligned} ||\boldsymbol{J}_b - \boldsymbol{J}^*||_e^2 \quad \leq \quad & (1-\phi\gamma)^{2b}||\boldsymbol{J}_0 - \boldsymbol{J}^*||_e^2 \\ & + \gamma^2 \sum_{i=0}^{b-1} ||\boldsymbol{\eta}_i||_e^2 (1-\phi\gamma)^{2(b-1-i)} \end{aligned}$$

Because $||\boldsymbol{\eta}_i||_e^2 \leq \sigma^2$ for all i we have:

$$\begin{aligned} ||\boldsymbol{J}_b - \boldsymbol{J}^*||_e^2 \quad \leq \quad & (1-\phi\gamma)^{2b}||\boldsymbol{J}_0 - \boldsymbol{J}^*||_e^2 \\ & + \gamma^2\sigma^2 \frac{1-(1-\phi\gamma)^{2b}}{1-(1-\phi\gamma)^2} \end{aligned}$$

Simplifying the above expression yields the result of part (b). Part (c) is an immediate consequence. □

We now show that an implementation that chooses $I(t)$ over a frame according to (32), using the $\boldsymbol{J}_b$ estimate instead of the optimal $\boldsymbol{J}^*$ vector, results in an approximation to the stochastic shortest path problem that deviates by an amount that depends on $||\boldsymbol{J}_b - \boldsymbol{J}^*||_e$.

Lemma 4: Suppose we choose $I(t)$ according to (32) over the course of a frame, using a vector $\boldsymbol{J}$ rather than $\boldsymbol{J}^*$. Let $\tilde{\boldsymbol{J}}(\boldsymbol{J})$ represent the expected sum cost over the frame (given $\boldsymbol{J}$). Then:

$$||\tilde{\boldsymbol{J}}(\boldsymbol{J}) - \boldsymbol{J}^*||_e \leq \frac{2(1-\phi)||\boldsymbol{J} - \boldsymbol{J}^*||_e}{\phi} \tag{89}$$

Further, defining $\mathbf{1}$ as a vector of all 1 entries with size equal to the dimension of $\boldsymbol{J}$, we have:

$$\mathbb{E}\left\{\tilde{\boldsymbol{J}}(\boldsymbol{J})\right\} \leq \boldsymbol{J}^* + \mathbf{1}\frac{2(1-\phi)||\boldsymbol{J} - \boldsymbol{J}^*||_e}{\phi}$$

where the expectation above is with respect to the randomness of the $\boldsymbol{J}$ vector.

Proof: Let $I(t)$ represent the control decision on slot t made using the $\boldsymbol{J}$ vector, and let $I^*(t)$ represent the decision that would be made under the $\boldsymbol{J}^*$ vector. Then:

$$\begin{aligned}\tilde{\boldsymbol{J}}(\boldsymbol{J}) = \phi\mathbb{E}&\left\{\min_{I\in\mathcal{I}_{[\omega(t),1],\boldsymbol{z}}} \boldsymbol{c}^{(1)}_{\boldsymbol{\Theta}}(I,\omega(t))\right\}\\ &+(1-\phi)\mathbb{E}\left\{\boldsymbol{c}^{(0)}_{\boldsymbol{\Theta}}(I(t),\omega(t)) + P^{(0)}(I(t),\omega(t))\tilde{\boldsymbol{J}}(\boldsymbol{J}) \mid \boldsymbol{J}\right\}\end{aligned}$$

where the expectation is with respect to the random $\omega(t)$ outcome. Thus:

$$\begin{aligned}\tilde{\boldsymbol{J}}(\boldsymbol{J}) = \phi\mathbb{E}&\left\{\min_{I\in\mathcal{I}_{[\omega(t),1],\boldsymbol{z}}} \boldsymbol{c}^{(1)}_{\boldsymbol{\Theta}}(I,\omega(t))\right\}\\ &+(1-\phi)\mathbb{E}\left\{\boldsymbol{c}^{(0)}_{\boldsymbol{\Theta}}(I(t),\omega(t)) + P^{(0)}(I(t),\omega(t))\boldsymbol{J} \mid \boldsymbol{J}\right\}\\ &+(1-\phi)\mathbb{E}\left\{P^{(0)}(I(t),\omega(t))\right\}(\tilde{\boldsymbol{J}}(\boldsymbol{J}) - \boldsymbol{J})\end{aligned} \tag{90}$$

Because $I(t)$ minimizes the second term of the above equality, we have:

$$\begin{aligned}(1-\phi)&\mathbb{E}\left\{\boldsymbol{c}^{(0)}_{\boldsymbol{\Theta}}(I(t),\omega(t)) + P^{(0)}(I(t),\omega(t))\boldsymbol{J} \mid \boldsymbol{J}\right\}\\ &\leq (1-\phi)\mathbb{E}\left\{\boldsymbol{c}^{(0)}_{\boldsymbol{\Theta}}(I^*(t),\omega(t)) + P^{(0)}(I^*(t),\omega(t))\boldsymbol{J} \mid \boldsymbol{J}\right\}\\ &= (1-\phi)\mathbb{E}\left\{\boldsymbol{c}^{(0)}_{\boldsymbol{\Theta}}(I^*(t),\omega(t)) + P^{(0)}(I^*(t),\omega(t))\boldsymbol{J}^* \mid \boldsymbol{J}\right\}\\ &\quad (1-\phi)\mathbb{E}\left\{P^{(0)}(I^*(t),\omega(t))(\boldsymbol{J} - \boldsymbol{J}^*) \mid \boldsymbol{J}\right\}\end{aligned}$$

Combining the above with (90) yields:

$$\begin{aligned}\tilde{\boldsymbol{J}}(\boldsymbol{J}) &\leq \boldsymbol{J}^* + (1-\phi)\mathbb{E}\left\{P^{(0)}(I(t),\omega(t))\right\}(\tilde{\boldsymbol{J}}(\boldsymbol{J}) - \boldsymbol{J})\\ &+(1-\phi)\mathbb{E}\left\{P^{(0)}(I^*(t),\omega(t))\right\}(\boldsymbol{J} - \boldsymbol{J}^*)\end{aligned}$$

However, we also know that $\boldsymbol{J}^* \leq \tilde{\boldsymbol{J}}(\boldsymbol{J})$. Therefore, using the fact that the expectation of a transition matrix is also a transition matrix, and that $||P\boldsymbol{X}||_e \leq ||\boldsymbol{X}||_e$:

$$\begin{aligned}||\tilde{\boldsymbol{J}}(\boldsymbol{J}) - \boldsymbol{J}^*||_e &\leq (1-\phi)||\boldsymbol{J} - \boldsymbol{J}^*||_e + (1-\phi)||\tilde{\boldsymbol{J}}(\boldsymbol{J}) - \boldsymbol{J}||_e\\ &\leq (1-\phi)[||\boldsymbol{J} - \boldsymbol{J}^*||_e + ||\tilde{\boldsymbol{J}}(\boldsymbol{J}) - \boldsymbol{J}^*||_e + ||\boldsymbol{J}^* - \boldsymbol{J}||_e]\end{aligned}$$

Rearranging terms yields (89).

To prove the final part of the lemma, note that for each entry i we have:

$$\tilde{J}(\boldsymbol{J})[i] \leq J^*[i] + |\tilde{J}(\boldsymbol{J})[i] - J^*[i]|$$

and hence (because $\mathbb{E}\left\{|X|\right\} \leq \sqrt{\mathbb{E}\left\{X^2\right\}}$ for any random variable X):

$$\mathbb{E}\left\{\tilde{J}(\boldsymbol{J})[i]\right\} \leq J^*[i] + ||\tilde{\boldsymbol{J}}(\boldsymbol{J}) - \boldsymbol{J}^*||_e$$

□

Appendix F

Here we state a simple and well known bound on the variance of a bounded random variable. The proof is provided for completeness.

Lemma 11: (Maximum Variance of a Bounded Random Variable) Let X be a random variable such that $x_{min} \leq X \leq x_{max}$ with probability 1, where x_{min} and x_{max} are finite constants. Then:

(a) The variance, denoted by $Var(X)$, is finite and:

$$Var(X) \leq \frac{(x_{max} - x_{min})^2}{4} \tag{91}$$

Further, the above bound is tight and is achieved by the following extremal distribution:

$$Pr[X = x_{min}] = Pr[X = x_{max}] = 1/2$$

(b) If X additionally has a known mean $\overline{X}$, then the variance satisfies the following tighter constraint:

$$Var(X) \leq (x_{max} - \overline{X})(\overline{X} - x_{min}) \tag{92}$$

Further, the above bound is tight and is achieved by the following extremal distribution:

$$\begin{aligned}Pr[X = x_{min}] &= \frac{x_{max} - \overline{X}}{x_{max} - x_{min}}\\ Pr[X = x_{max}] &= \frac{\overline{X} - x_{min}}{x_{max} - x_{min}}\end{aligned}$$

Note that in the special case when $\overline{X} = 0$, the above lemma implies $Var(X) \leq |x_{max}x_{min}|$.

Proof: (Lemma 11) We have:

$$\begin{aligned}Var(X) &= Var(X - x_{min})\\ &= \mathbb{E}\left\{(X - x_{min})^2\right\} - (\overline{X} - x_{min})^2\\ &\leq \mathbb{E}\left\{(x_{max} - x_{min})(X - x_{min})\right\} - (\overline{X} - x_{min})^2\\ &= (x_{max} - x_{min})(\overline{X} - x_{min}) - (\overline{X} - x_{min})^2\\ &= (\overline{X} - x_{min})(x_{max} - \overline{X})\end{aligned}$$

where the inequality in the above chain of expressions follows because $0 \leq (X - x_{min}) \leq (x_{max} - x_{min})$ with probability 1, and so $(X - x_{min})^2 \leq (x_{max} - x_{min})(X - x_{min})$ with probability 1. This proves (92) in part (b).

To prove (91) in part (a), we have:

$$\begin{aligned}Var(X) &\leq (\overline{X} - x_{min})(x_{max} - \overline{X})\\ &\leq \max_{x_{min}\leq x\leq x_{max}} (x - x_{min})(x_{max} - x)\end{aligned}$$

where the final inequality follows because $x_{min} \leq \overline{X} \leq x_{max}$. By taking a derivative with respect to x, it is easy to show that:

$$\max_{x_{min}\leq x\leq x_{max}} (x - x_{min})(x_{max} - x) = \frac{(x_{max} - x_{min})^2}{4}$$

This proves (91). That the bounds (91) and (92) are achieved at the given extremal distributions is easily verified. □

Appendix G – Proof of Lemma 5

Here we prove Lemma 5 of Section IV-D, restated below for convenience.

Lemma 5: For the vectors $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$, and for the β value defined in (43), we have:

(a) The difference between $\boldsymbol{J}_{\boldsymbol{\Theta}_1}$ and $\boldsymbol{J}_{\boldsymbol{\Theta}_2}$ satisfies:

$$||\boldsymbol{J}_{\boldsymbol{\Theta}_1} - \boldsymbol{J}_{\boldsymbol{\Theta}_2}||_d \leq \frac{\beta}{\phi}$$

(b) Let $I_1(t)$ denote the policy decisions at time t under the policy that makes optimal decisions subject to queue backlogs $\boldsymbol{\Theta}_1$, and define $\boldsymbol{J}_{21}^{mis}$ as the expected sum cost over a frame of a mismatched policy *that incurs costs according to backlog vector $\boldsymbol{\Theta}_2$ but makes decisions according to $I_1(t)$ (and hence has the same frame duration and decisions as the optimal policy for $\boldsymbol{\Theta}_1$). Then:*

$$\boldsymbol{J}_{\boldsymbol{\Theta}_2} \leq \boldsymbol{J}_{21}^{mis} \leq \boldsymbol{J}_{\boldsymbol{\Theta}_1} + \boldsymbol{1}\frac{\beta}{\phi}$$

where $\boldsymbol{1}$ is a vector of all 1 values with the same dimension as $\boldsymbol{J}_{\boldsymbol{\Theta}_1}$.

Proof: By definition, we have $\boldsymbol{J}_{\boldsymbol{\Theta}_2} \leq \boldsymbol{J}_{21}^{mis}$ (as $\boldsymbol{J}_{\boldsymbol{\Theta}_2}$ is the minimum sum cost over any policy when penalties are incurred according to $\boldsymbol{\Theta}_2$ queue backlog). Consider any entry z, and suppose we start in initial state $z(0) = z$.[7] Let T_1 denote the renewal time under policy $I_1(t)$, and let $z_1(\tau)$ denote the state at time τ under policy $I_1(t)$. Then:

$$\begin{aligned}
J_{\boldsymbol{\Theta}_2}[z] &\leq J_{21}^{mis}[z] \\
&= \mathbb{E}\left\{\sum_{\tau=0}^{T_1-1} c_{\boldsymbol{\Theta}_2}(I_1(\tau), \Omega(\tau), z_1(\tau))\right\} \\
&= J_{\boldsymbol{\Theta}_1}[z] + \mathbb{E}\left\{\sum_{\tau=0}^{T_1-1} c_{\boldsymbol{\Theta}_2}(I_1(\tau), \Omega(\tau), z_1(\tau))\right\} \\
&\quad -\mathbb{E}\left\{\sum_{\tau=0}^{T_1-1} c_{\boldsymbol{\Theta}_1}(I_1(\tau), \Omega(\tau), z_1(\tau))\right\} \\
&\leq J_{\boldsymbol{\Theta}_1}[z] + \frac{\beta}{\phi}
\end{aligned}$$

where the final inequality is due to the fact that the mean renewal time is at most $1/\phi$. This proves part (b).

To prove part (a), note that part (b) implies:

$$\boldsymbol{J}_{\boldsymbol{\Theta}_2} \leq \boldsymbol{J}_{\boldsymbol{\Theta}_1} + \boldsymbol{1}\frac{\beta}{\phi}$$

However, switching the roles of $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$, we can similarly derive $\boldsymbol{J}_{\boldsymbol{\Theta}_1} \leq \boldsymbol{J}_{\boldsymbol{\Theta}_2} + \boldsymbol{1}\beta/\phi$. This proves part (a). □

References

[1] L. Georgiadis, M. J. Neely, and L. Tassiulas. Resource allocation and cross-layer control in wireless networks. *Foundations and Trends in Networking*, vol. 1, no. 1, pp. 1-149, 2006.

[2] D. P. Bertsekas and J. N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, Belmont, Mass, 1996.

[3] S. Ross. *Introduction to Probability Models*. Academic Press, 8th edition, Dec. 2002.

[4] E. Altman. *Constrained Markov Decision Processes*. Boca Raton, FL, Chapman and Hall/CRC Press, 1999.

[5] S. Meyn. *Control Techniques for Complex Networks*. Cambridge University Press, 2008.

[6] J. Abounadi, D. Bertsekas, and V. S. Borkar. Learning algorithms for markov decision processes with average cost. *SIAM Journal on Control and Optimization*, vol. 20, pp. 681-698, 2001.

[7] L. Tassiulas and A. Ephremides. Dynamic server allocation to parallel queues with randomly varying connectivity. *IEEE Transactions on Information Theory*, vol. 39, pp. 466-478, March 1993.

[8] E. M. Yeh. *Multiaccess and Fading in Communication Networks*. PhD thesis, Massachusetts Institute of Technology, Laboratory for Information and Decision Systems (LIDS), 2001.

[9] A. Ganti, E. Modiano, and J. N. Tsitsiklis. Optimal transmission scheduling in symmetric communication models with intermittent connectivity. *IEEE Transactions on Information Theory*, vol. 53, no. 3, March 2007.

[10] A. Fu, E. Modiano, and J. Tsitsiklis. Optimal energy allocation for delay-constrained data transmission over a time-varying channel. *Proc. IEEE INFOCOM*, 2003.

[11] M. Goyal, A. Kumar, and V. Sharma. Power constrained and delay optimal policies for scheduling transmission over a fading channel. *Proc. IEEE INFOCOM*, April 2003.

[12] N. Salodkar, A. Bhorkar, A. Karandikar, and V. S. Borkar. An on-line learning algorithm for energy efficient delay constrained scheduling over a fading channel. *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 732-742, May 2008.

[13] D. V. Djonin and V. Krishnamurthy. q-learning algorithms for constrained markov decision processes with randomized monotone policies: Application to mimo transmission control. *IEEE Transactions on Signal Processing*, vol. 55, no. 5, May 2007.

[14] R. Berry and R. Gallager. Communication over fading channels with delay constraints. *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135-1149, May 2002.

[15] M. J. Neely. Optimal energy and delay tradeoffs for multi-user wireless downlinks. *IEEE Transactions on Information Theory*, vol. 53, no. 9, pp. 3095-3113, Sept. 2007.

[16] M. J. Neely. Super-fast delay tradeoffs for utility optimal fair scheduling in wireless networks. *IEEE Journal on Selected Areas in Communications, Special Issue on Nonlinear Optimization of Communication Systems*, vol. 24, no. 8, pp. 1489-1501, Aug. 2006.

[17] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, 1995.

[18] M. J. Neely. Energy optimal control for time varying wireless networks. *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 2915-2934, July 2006.

[19] A. Stolyar. Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm. *Queueing Systems*, vol. 50, pp. 401-457, 2005.

[20] A. Stolyar. Greedy primal-dual algorithm for dynamic resource allocation in complex networks. *Queueing Systems*, vol. 54, pp. 203-220, 2006.

[21] M. J. Neely, E. Modiano, and C. Li. Fairness and optimal stochastic control for heterogeneous networks. *Proc. IEEE INFOCOM*, March 2005.

[22] D. P. Bertsekas and R. Gallager. *Data Networks*. New Jersey: Prentice-Hall, Inc., 1992.

[23] R. Gallager. *Discrete Stochastic Processes*. Kluwer Academic Publishers, Boston, 1996.

[24] M. J. Neely. *Dynamic Power Allocation and Routing for Satellite and Wireless Networks with Time Varying Channels*. PhD thesis, Massachusetts Institute of Technology, LIDS, 2003.

[25] M. J. Neely. Distributed and secure computation of convex programs over a network of connected processors. *DCDIS Conf., Guelph, Ontario*, July 2005.

[26] D. P. Bertsekas. *Dynamic Programming and Optimal Control, vols. 1 and 2*. Athena Scientific, Belmont, Mass, 1995.

[27] M. J. Neely. Max weight learning algorithms with application to scheduling in unknown environments. *arXiv:0902.0630v1*, Feb. 2009.

[28] M. J. Neely, E. Modiano, and C. E Rohrs. Dynamic power allocation and routing for time varying wireless networks. *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 89-103, January 2005.

[7] Note that while all frames start with $z = 0$, and hence have optimal cost $J^*[0]$, $\boldsymbol{J}^*$ is defined with entries indexed by general initial states z.